

# SEARCH REQUEST FORM

Access DB# \_\_\_\_\_

Scientific and Technical Information Center

Requester's Full Name: \_\_\_\_\_ Examiner #: \_\_\_\_\_ Date: \_\_\_\_\_  
 Art Unit: \_\_\_\_\_ Phone Number 30 \_\_\_\_\_ Serial Number: \_\_\_\_\_  
 Mail Box and Bldg/Room Location: \_\_\_\_\_ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

\*\*\*\*\*  
 Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc. if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: \_\_\_\_\_

Inventors (please provide full names): \_\_\_\_\_

Earliest Priority Filing Date: \_\_\_\_\_

*\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

## STAFF USE ONLY

Searcher: _____	Type of Search	Vendors and cost where applicable
Searcher Phone #: _____	NA Sequence (#) <u>3</u>	STN _____
Searcher Location: _____	AA Sequence (#) <u>1</u>	Dialog _____
Date Searcher Picked Up: <u>10/17/01</u>	Structure (#) _____	Questel/Orbi: _____
Date Completed: <u>10/18/01</u>	Bibliographic _____	Dr Link _____
Searcher Prep & Review Time _____	Litigation _____	Lexis/Nexis _____
Clencal Prep Time _____	Fulltext _____	Sequence Systems <u>\$3</u>
Online Time: _____	Patent Family _____	WWW/Internet _____
	Other _____	Other (specify) _____



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2001, 14:43:25 ; Search time 9769.47 Seconds

(without alignments)  
5069.644 Million cell updates/sec

Title: US-09-502-426-1\_COPY\_1\_3202  
Perfect score: 3202

Sequence: 1 atgtgggtatataatgttg.....gagagagagaactagcttc 3202

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_ba3:\*  
4: gb\_in1:\*  
5: gb\_in2:\*  
6: gb\_in3:\*  
7: gb\_om:\*  
8: gb\_ov:\*  
9: gb\_pat1:\*  
10: gb\_pat2:\*  
11: gb\_ph:\*  
12: gb\_pl1:\*  
13: gb\_pl2:\*  
14: gb\_pl3:\*  
15: gb\_pl4:\*  
16: em\_ba1:\*  
17: em\_ba2:\*  
18: em\_fun:\*  
19: em\_htgo\_hum:\*  
20: em\_htgo\_inv:\*  
21: em\_htgo\_rod:\*  
22: em\_htg\_hum1:\*  
23: em\_htg\_hum2:\*  
24: em\_htg\_hum3:\*  
25: em\_htg\_hum4:\*  
26: em\_htg\_hum5:\*  
27: em\_htg\_hum6:\*  
28: em\_htg\_hum7:\*  
29: em\_htg\_hum8:\*  
30: em\_htg\_inv1:\*  
31: em\_htg\_inv2:\*  
32: em\_htg\_other:\*  
33: em\_htg\_rod:\*  
34: em\_hum1:\*  
35: em\_hum2:\*  
36: em\_hum3:\*  
37: em\_hum4:\*  
38: em\_hum5:\*  
39: em\_hum6:\*  
40: em\_hum7:\*  
41: em\_in:\*  
42: em\_om:\*  
43: em\_or:\*

44: em\_ov:\*  
45: em\_pat:\*  
46: em\_ph:\*  
47: em\_pl:\*  
48: em\_ro:\*  
49: em\_sts:\*  
50: em\_sy:\*  
51: em\_un:\*  
52: em\_v1:\*  
53: gb\_sts1:\*  
54: gb\_sts2:\*  
55: gb\_sts3:\*  
56: gb\_sy:\*  
57: gb\_un:\*  
58: gb\_v11:\*  
59: gb\_v12:\*  
60: gb\_htg1:\*  
61: gb\_htg2:\*  
62: gb\_htg3:\*  
63: gb\_htg4:\*  
64: gb\_htg5:\*  
65: gb\_htg6:\*  
66: gb\_htg7:\*  
67: gb\_htg8:\*  
68: gb\_htg9:\*  
69: gb\_htg10:\*  
70: gb\_htg11:\*  
71: gb\_htg12:\*  
72: gb\_htg13:\*  
73: gb\_htg14:\*  
74: gb\_htg15:\*  
75: gb\_htg16:\*  
76: gb\_htg17:\*  
77: gb\_htg18:\*  
78: gb\_htg19:\*  
79: gb\_htg20:\*  
80: gb\_htg21:\*  
81: gb\_htg22:\*  
82: gb\_htg23:\*  
83: gb\_htg24:\*  
84: gb\_htg25:\*  
85: gb\_pr1:\*  
86: gb\_pr2:\*  
87: gb\_pr3:\*  
88: gb\_pr4:\*  
89: gb\_pr5:\*  
90: gb\_pr6:\*  
91: gb\_pr7:\*  
92: gb\_pr8:\*  
93: gb\_pr9:\*  
94: gb\_ro1:\*  
95: gb\_ro2:\*  
96: gb\_in4:\*  
97: gb\_pr10:\*  
98: em\_ba3:\*

Pred. NO. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	3134.2	97.9	84196	14	ATT3A5	AL132979 Arabidops
2	1132	35.4	4818	12	AF044216	AF044216 Arabidops
3	133.2	4.2	104992	60	AC005504	AC005504 Plasmodu
4	133.2	4.2	169546	60	AC004157	AC004157 Plasmodu
C 5	130.8	4.1	104992	60	AC005504	AC005504 Plasmodu
C 6	130.8	4.1	169546	60	AC004157	AC004157 Plasmodu
7	129.4	4.0	14867	4	AE001398	AE001398 Plasmodu
8	127.8	4.0	86827	96	PFMAL3P5	AL034556 Plasmodu

```

source          1. 84196
                /organism="Arabidopsis thaliana"
                /variety="Columbia"
                /db_xref="taxon:3702"
                /chromosome="3"
                2983, 3338
                /note="355 bp LTR"
                3339..7564
                /note="LTR-transposon"
                7565..7922
                /note="367bp LTR"
                12339..13197
                /gene="T3A5.10"
                join(12339..12665,12742..12803,12888..13055,13125..13197)
                /gene="T3A5.10"
                /note="similarity to cyclin-dependent kinase inhibitor
                (ICK1), Arabidopsis thaliana, EMBL:ATU94772.1"
                /codon_start=1
                /product="putative protein"
                /protein_id="CAB62432.1"
                /db_xref="GI:6561966"
                /translation="MAAARRRRRDVVEENGVTTTTVKRRKMEEVLDVESHILISPCV
                QATNNGGVARNSSAGASSTSYIVYIRRRDSPVEEQCOGLEEDSSVSCSTSESKRR
                IEEVDLEENDDRETETSWLYDDLNKSEESMNNDSVAIVEDESRRRLRLSHETV
                KEALDEDFQVAEKDLRNLKLLCSMKYNFDEKDEPLQGRYEWKLNIP"
                12339..12665
                /gene="T3A5.10"
                /number=1
                12666..12741
                /gene="T3A5.10"
                /number=1
                12742..12803
                /gene="T3A5.10"
                /number=2
                12804..12887
                /gene="T3A5.10"
                /number=2

```

exon /number=3  
 13125..13197  
 /gene="T3A5.10"  
 /number=4  
 complement(15858..15901)  
 /number=1  
 15858..16470  
 /gene="T3A5.20"  
 complement(join(15858..15901,16014..16470))  
 /note="similarity to predicted proteins, Arabidopsis  
 thaliana"  
 /codon\_start=1  
 /product="putative protein"  
 /protein\_id="CAB62433.1"  
 /db\_xref="GI:5561967"  
 /translation="MARLEHNSMKGCAARISFSNEVEINSEKSNMKNINSSTSS  
 MPSADFAVTDYMSIPADEIFLGKGIIPKETSHTVHTLGEELLTEEGSMVGNTF  
 SLRILLSSSSFTSGTWRRLGLKRTHVRSKKTDKNEEVLISODHRIISGVATREC  
 QVADPTR"  
 complement(15902..16013)  
 /number=1  
 complement(16014..16470)  
 /number=2  
 18303..19931  
 /gene="T3A5.30"  
 complement(18303..19931)  
 /number=1  
 complement(18303..19931)  
 /codon\_start=1  
 /product="scarecrow-like 7 (SCL7)"  
 /protein\_id="CAB62434.1"  
 CDS  
 exon  
 intron



/db\_xref="GI:6561968"  
/translation="MAYMTDSDGNLMAIAQOLIKQKQOQOQSOHQOQEOEDQEPNPMWPN  
PSFTSLIGSGDFDQVNTDNGFHFPHLEHONAAVAASEEDSDMEMESLNGSDAS  
OTNPDEPIYHDDPFPSPRLSAPSYLRNVKNDSDASOOLPPPPASTATWMSPPSPQ  
HPPPPQPDNDLNPFIKAIHDVARKPEPTDPTLIRIKESSESQDPIORGYFPAR  
ALSHKETSPPSSSSSSLEDFLTKTINDACPKFAHLTNQALILEATONSNIHI  
VDFGLFQGIQWALLQALATRSKGPTRIRISGIPASLGSPGSLATGNRLDFA  
AILDNFEFFPYLPYLIQLLNGSSFPVDEVLVVMFELYLILDETAATVGTALRLA  
RSLNRIYVLGCEYSLNRVEFANKNLSREYSAVPESELEPNLDRDQERLRERVL  
GRRIMDLVRSDDNKPKGTGRGLMEKKQEGFEPVKRSNNAVSAQAKILLM  
NNYSTILSVSESGFISLANNVPLLTYSNR"  
26093..28999  
/gene="T3A5.40"  
/complement(join(26093..26338,27091..27200,27308..27386,  
27478..27570,27723..28037,28136..28288,28374..28698,  
28779..28999))  
/note="Contains Cytochrome P450 cysteine heme-iron ligand  
signature AA455-464;Prokaryotic membrane lipoprotein lipid  
attachment site AA320-330"  
/codon\_start=1  
/product="steroid 22-alpha-hydroxylase (DWF4)"  
/protein\_id="CAB62435.1"  
/db\_xref="GI:6561969"  
/translation="MFEETHLLPLLLPSLLLELLLELLKRRNRKTRPNLPQGS  
GMPLEGTIGLKPYTATTLGDGMQOHVSKYKIRSNLFEPTIVSADAGLRILQ  
NEGRLECSYSPRSIGILKWSMLVGDMDHMSISLNFSLHARLRLILKDYERH  
TLFVLDWQOQSIIFSADQEAKEFTNLMAKHMSMDPEEEOLEKKEVTPMKGVN  
APLNPTAVHRAKGLSRATILKFIERKMEERLDIKEDQEEEEKTEDEAESDSDH  
VKORTDDDLGKLVKHSNSTEQLDILSLSPGHETSSVAIALAIFLOACRKY  
ELREHELEIARAKKELGSELNMDYKMDTQCVINTELRGLNVRLHKKALKDV  
RKGVDIPSGWKLVPVISAVALHDSNRYPQPNLFWPWOQNNQSSSGSGSFTWGN  
NMPGGGPRLCAGSELAKLEMAVFIHHLVLFKNMELADEDDKPFAPFPDFPENGDIR  
VSRIL"  
complement(26093..26338)  
exon /number=1  
intron /complement(26339..27090)  
exon /number=1  
intron /complement(27091..27200)  
exon /number=2  
intron /complement(27201..27307)  
exon /number=2  
intron /complement(27308..27386)  
exon /number=3  
intron /complement(27387..27477)  
exon /number=3  
intron /complement(27478..27570)  
exon /number=4  
intron /complement(27571..27722)  
exon /number=4  
intron /complement(27723..28037)  
exon /number=5  
intron /complement(28038..28135)  
exon /number=5  
intron /complement(28136..28288)  
exon /number=6  
intron /complement(28289..28373)  
exon /number=6  
intron /complement(28374..28698)  
exon /number=7  
intron /complement(28699..28778)  
exon /number=7  
intron /complement(28779..28999)  
gene /number=8  
38307..41323  
/gene="T3A5.50"  
/complement(38307..38796)  
CDS /number=1  
complement(join(38307..38796,38877..39090,39189..39270,  
40180..40242,40337..40401,40489..40564,40655..40734,  
40805..40919,41225..41323))  
/note="Contains Eukaryotic putative RNA-binding region  
RNP-1 signature AAI79-186"  
/codon\_start=1

/product="U1 snRNP 70K protein"  
/protein\_id="CAB62436.1"  
/db\_xref="GI:6561970"  
/translation="MGDSGDPELRPMNAVOARAKVONRANYLQKLMGOSHPGLTN  
NLKLFPERPLEPEYPPPEKRCPPYTGMAOVSNFARPGDEYAPRPREVLPQKR  
EKIRHLREKGVKAEEDLKTDPNNDPATDPTKTLFVSLNLTNSESSESKIKREFS  
YPMRPRRGGGLGTSRVGGEIEIYEMQPOGRTSOSSEPREREKREKERE  
RSRELSDHPPRRSRDRPREDKHHRDROGGRDRDRDRDRDRDRDRDRDRDR  
GDRPSRDRDRSRKREDYDGEYEHGGGRSRDRDRDRDRDRDRDRDRDRDR  
TDYSHRDKMEEDFRTEREYKRSKRSREYVR"  
complement(38797..38876)  
intron /number=1  
exon /complement(38877..39090)  
exon /number=2  
intron /complement(39091..39188)  
exon /number=2  
intron /complement(39189..39270)  
exon /number=3  
intron /complement(39271..40179)  
exon /number=3  
intron /complement(40180..40242)  
exon /number=4  
intron /complement(40243..40336)  
Query Match 97.9%; Score 3134.2; DB 14; Length 84196;  
Best local similarity 99.5%; Pred No. 0;  
Matches 3166; Conservative 0; Mismatches 13; Indels 4; Gaps 4;  
QY 1 atgtggtatataatctgttgggtcgttgaagctacaataaattcgtgttctgt 60  
DB 32199 ATGTGGTATTATATATTGTTGGTTCGTTGACCTACAAATATAATTTCTGTTCTGGT 32140  
QY 61 tattcgttcacatgattgagtttggttcacaaatttggaatccaagataataatc 120  
DB 32139 TATTCTGTTCACTGATTGAGTTGGTTCCTCAATTGGATTCAGATTAATTAATTT 32080  
QY 121 aaaaatcattaaatattacagaatattacattacattgattgattataacaa 180  
DB 32079 AAAATTCATTTAAATATTACAGTAATTAAATTAATTAATTAATTAATTAATTA 32020  
QY 181 aatatcatcttgtgatalatgagaataatlgagtttgaattataataaagaagaa 240  
DB 32019 AATATCATCTTTGCTATATAGAGAAATATGAGCTTTGGAATTTATATATTAAGGAAA 31960  
QY 241 taaatgattccattcgttggatlaacagtaagttttggttccttcgttatacgt 300  
DB 31959 TAATCGATTCATTGTTGGATTACACAGTTAAAGTTTGTGTTCTTTGTTATATGT 31900  
QY 301 atatgagtaataaagaagtttgtaagttgaagtttaaacatttcgttataacccca 360  
DB 31899 ATATGAGTAATCAAAAAGCAATGTGTAAGTAAACATATTTCTGTATGACCCCC 31840  
QY 361 aaaaaaaaaaaaaaacaacaaaccccccccatagtttctgttcgtgatt 420  
DB 31839 CAAAAAATAAATAAACAACAAACAAA-ACAATTTTCGATATAGTTTGTGTTCTGTGATT 31781  
QY 421 aggttaattgataataatcatgatacttccttgattactatgaagaatttcctac 480  
DB 31780 AGGTTTATTTGATCATTAATACATGCACTATTCTTGTGATTACTATGAAATTTCTTAC 31721  
QY 481 caatataaatttcgaattcatctcttggttataataataatagaagtgataatcc 540  
DB 31720 CAATTAATAATTTGAAATTCATCTCTTGTATTAATTAATTAATTAATTAATTAATTC 31661  
QY 541 gttatcgaatccatccaatatgataatgattcttgytcaatccagcaattataa 600  
DB 31660 GTTATTCGATCACCATCATATGATTAATGATTTCTTGCTGAATTCAGCAATTAATTAACA 31601  
QY 601 agagttatgagaataaaccgaataaagaagaagtagtgacccatgagatag 660  
DB 31600 AGAGTATTGAGAAAAAAGCAAAATTAAGAAAGCAAGAGTAGTACCCTAGAGATATG 31541

QY	661	tgaaacaaatccaaagagaataagagatgacaaaccaaagggttggaataatgylccct	720
Db	31540	TCGAATATATATCAAGAGAGATAGAGANTGACACCAACAAAGGTTGGATATATGTCCT	31481
QY	721	ggcagcttcctccaaacaaatcagaccatttggattcttcggatcattcgttaaa	780
Db	31480	GCCAGCTTCTCTCCAAATCAATATCGACCCTATTTGGATTTTCTGCGATTCGAATTCGTTAANA	31421
QY	781	tttcggataacgatttggaaaaatatlttatltgttagctgtaacctaatatgtlcca	840
Db	31420	TTTTGGCATACCATGTGTGAAAAATATTTTATTTGTTAGCTGATCTCAATATATGTCCA	31361
QY	841	ggtattttagataacttcgtttaaagatatgtgtcttcttcttltgttgcgttctc	900
Db	31360	GGATTTTGGATATATCTCTGTTTAAAGCATATTTGTCTCTTTTGTGTTGGTTCTCT	31301
QY	901	taacatacatatcgcgagatacatgatbaacaatgatataatcacaaacaaatgtctg	960
Db	31300	TAACTATATATATATGCGCGATATATGATATACATGATATATCCAAACAAATATGTCGGG	31241
QY	961	accattttgataaaacttttctcaaacatctacgggacatggaactcgaccttaata	1020
Db	31240	ACCATTTTGGAAATAACTTTTCTCAAACTTCGGGACACTGACCTCGAACCTT-AAATA	31182
QY	1021	cgatttcaagcgtcacagctgtgagatcctgacataagacaataagacccgttcaac	1080
Db	31181	CGATTTTACGGGTCACATGATTGAGATTATCTAGCATTAAGCATTAAGACCCGTTCAACG	31122
QY	1081	tatttatacaagttaccaaacctgaatatagcttgaactcccttagaaaaatttgaatha	1140
Db	31121	TATTTATACAAAGTTACAAACGAATATATGCTTGAAATCTTTACAAAAATTTTGGATTA	31062
QY	1141	ccggtctgattatgaatatagatttagtbgtaacaacaatgttaactaatagtgtca	1200
Db	31061	CCGGTTGTTATGTAATATATATAGTTTAGTGTGTAACAAATATGTTAATCAATTAGTGTGCA	31002
QY	1201	acatatacaatactcccttaacagaaaaaacaacttaaggaagttaacatatccatat	1260
Db	31001	ACATATACATTAATTCCTTACACAAAAAACMACTTAAGAGAGTTTACATATCCATATAT	30942
QY	1261	gggtatgcatcaccttccacgtcatctgatactagaagacaagaagaatagtatgtagtc	1320
Db	30941	GGGTATGCTATACCTTTCACGATGTCTATACTAGAGACTTAAGAATAGTTATGTATGTCTC	30882
QY	1321	gataaaatgaatctcacacgcgttggtaataatatayggaaccgtatgtltaacgactgcga	1380
Db	30881	GATTAATATGAATTTCCACGCGCGTGAATTAATATGTGGACCGTATGTTACGATCCTGCNA	30822
QY	1381	atatatcttcgtgttgcacaacataaaaaacaagaaaaaagaagaacgatttct	1440
Db	30821	ATATATTTCTTGTTGGTGCACAAATTAATAAACAACAAACAGAAAAAGAAAAAGCAATTTT	30762
QY	1441	cttggatccatcaatgatcctcaaaatgcatagatccttggctacagtttcgaagctc	1500
Db	30761	CTTGATTTCCATTCATGATGATCTTAATATGATGATCTTTTGGTTACAGATTTCGAAGTCC	30702
QY	1501	ctacaacgcgttgaaccatctgcaacatlaaatgtcttcccttaatgcatcttaacat	1560
Db	30701	TCTACAGCGTGTACCACTGTCCACTATTAATATGCTTTCTTTATGCACTTAATACAT	30642
QY	1561	atttatgttagtggaaatataaaggcgaacttgaacgttaacgtataattatattag	1620
Db	30641	ATTATTTGTTAGTTGAATTTTAATTAAGCGCAACTTGTACATTAACAATATTTAATATTAG	30582
QY	1621	atacagatagtgatctatcccaatataacttggatgtttaaactaatctgttct	1680
Db	30581	ATACATGATATGATTTATTCCAAAATACATACTTTGATGTTTAAACTTAATCTGTTTCT	30522
QY	1681	tcctcaggtataaatataatcatccgaggtaaaaagtttgccttaatttcgcgatgc	1740
Db	30521	TCCTACG-CTATTAATAATTAATCATCTGAGGTAAAAAAGTTTGTCTTTATTTCCGGAATGC	30463

QY	1741	atgaaggaataacctaatgacttaattcttttgaaaaatgaacctctttactccatagatc	1800
Db	30462	atgaaggatataacctaatgacttaattcttttgaaaaatgaacctctttactccatagatc	30400
QY	1801	aattccgatctgtttctgttgccataatgcagccctcccaactgtagatgcaattcttc	18600
Db	30402	aatttccgatctgtttctgttgccataatgcagccctcttccaaactgtagatgcaattcttc	30340
QY	1861	ctcgcaaalatalaaattaggaaatccaatgctactacataagaagaacagctgtagatc	19200
Db	30342	tctgcgaatatttaattataggaaatccaatgctactacataagaagaacagctgtagatc	30280
QY	1921	acattttaaatttaaagccaaaatttttgaaaaatgctataaattcttcaacatatataaa	19800
Db	30282	acattttaaatttaaaagccaaaatttttgaaaaatgctataaattcttcaacataattataa	30222
QY	1981	aataatgacctataatgatacttccctaagctctcttaaaatctttcttttataattagta	20400
Db	30222	aatatgacctatataatgatacttccctaagctctctttaaattctttcttttataattagta	30160
QY	2041	taaaatacatltagaaccataatagtttgttgaattccaatctccatcaatlatcttttg	21000
Db	30162	taaaatacatatttagaaccataatagtttgttgaattccaatctccatcaatlatcttttg	30100
QY	2101	aaactacataacttatataatttagtacaataatgcataagaagtctccaaaaaaatc	21600
Db	30102	aaactacataacttttataatttttagtacaataatgcataagaagtctccaaaaaaattt	30040
QY	2161	ttgtttaacagaactctccaattctttcttttataagaacagaataataacagataagaa	22200
Db	30042	ttgttttaacagaaactctccaatttttttttttttttttttttttttttttttttttttt	29980
QY	2221	ctattttgttgtaagatggaagttagtaataatacatgaagcaaatcttataaaatatata	22800
Db	29982	ctattttgttgtaagatggaagtagtaataatacatgaagcaaatcttataaaatatata	29920
QY	2281	agccatacagcgcccaagaatgatacttctctgtagtgaatataaagatgtagtgcgat	23400
Db	29922	agccatatacagcgcccaagaatgtagtgaatgtagtgaatataaagatgtagtgcgat	29860
QY	2341	tcagaatttggagacacaaatgaacacgaatctaaatcattaaacttaaaataaataaat	24000
Db	29862	tcagaatttggagacacacaaatgaacacgaatctaaatcattaaacttaaaataaataaat	29800
QY	2401	ttgagtaaatgttgtttctcgaactatgaggggcaaaaaaagacaaatgcacaaagtctac	24600
Db	29802	ttgagtaaatgttgtttctcgaactatgaggggcaaaaaaagacaaatgcacaaagtctac	29740
QY	2461	gggtttgtagctgtccagttctggtataataatcaataaactcgtctttgacgcgacgtctg	25200
Db	29742	gggttttgcctgtccagttccgttatataatcaataaactcgtcttttgcacgcgacgtctg	29680
QY	2521	taggggtcccttcgaacttttccaactgtttccaccctacgcgtgagcccaaccttttccca	25800
Db	29682	taggggtcccttcgaacttttccaactgtttccaccctacgcgtgagcccaaccttttccca	29620
QY	2581	latccataagggtaattcttggaataatcccaatttaaacacgaattgagacgcttacccgactcc	26400
Db	29622	latccataagggtaatttttggaataatcccaatttttaaacacgattgagacgcttacccgactcc	29560
QY	2641	tggagatctcgtctgtagacattatcaaaaatttattagcagaaatgggttttatatttaa	27000
Db	29562	tggagattctcgtctgtagacattatcaaaaattttattagcagaaatgggttttatatttaa	29500
QY	2701	aactccaactttagatcagataaaaaatttcaataaacctttttagatggaattcgtcagatct	27600
Db	29502	aactccaactttagatcagataaaaaattttcaataaaccttttttagatggaattcgtcagatct	29440
QY	2761	atccaatgaactttttttcttcaacacgggtgtagaagaattatagttacattttagccagag	28200
Db	29442	atcttaatgaacttt	29380
QY	2821	acaattgattatagataataccaatttaaccatgataatttatgataataaataagctgtttaa	28800

|||||  
Db 29382 ACAATTGATTATAGATATATTCATTATTCATGATATTTATGATATTAATAGCTGTAA 29323  
Oy 2881 cttttccacccgcgcagctttgcacatttgcttttaatttaagaatttaataa 2940  
Db 29322 CTAATTCACATCGCGAGCTTTCGCACTTTGTTTAAATTAAGAGTTTAAATTA 29263  
Oy 2941 aaglttaaagagacataacagagcaacaaagtaataacacagagagaacaaagca 3000  
Db 29262 AAGTATTAAAGAGAGATACGAGGCAACAAAGTATACACGAGAGAAACAAAGCCA 29203  
Oy 3001 tgaagctcatggtta-gttaagcttaataagaagatttataatttaagcagat 3059  
Db 29202 TGAAGTCTCATGTTGTTAGTTTAAGCTTAATAAGAGATTATTAATTTAATGACAT 29143  
Oy 3060 gataacaatatatttttcgacttttaaacccctttcaacaacagaagctccctt 3119  
Db 29142 GATTAACATTTATTTTCTGACTCTTTTAAACCCCTCTTACAAACAGAGCTCCCTT 29083  
Oy 3120 ttcagtaagagtcgcagatcccaatcttaagaacaaagcattagaagaagaatgagtg 3179  
Db 29082 TTCAGTAGAAGATCCGATTCCTCAATCTTAAGACAAAGCCATTAGAAAGAGAGTGA 29023  
Oy 3180 agagagagagagaagaactagctcc 3202  
Db 29022 AGAGAGAGAGAGAACTAGCTCC 29000  
RESULT 2  
AF044216 4818 bp DNA PLN 06-MAR-1998  
LOCUS Arabidopsis thaliana steroid 22-alpha-hydroxylase (DMF4) gene,  
DEFINITION complete cds.  
ACCESSION AF044216  
VERSION AF044216.1 GI:2935341  
KEYWORDS  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
Brassicales; Brassicaceae; Arabidopsids.  
REFERENCE 1 (bases 1 to 4818)  
Choe,S., Dilkes,B.P., Fujioaka,S., Takatsuto,S., Sakurai,A. and  
Feldmann,K.A.  
TITLE The DMF4 gene of Arabidopsis encodes a cytochrome P450 that  
mediates multiple 22 alpha hydroxylation steps in brassinosteroid  
biosynthesis  
JOURNAL Plant Cell 10 (1998) In press  
REFERENCE 2 (bases 1 to 4818)  
Choe,S., Dilkes,B.P., Azpiroz,R. and Feldmann,K.A.  
TITLE Direct Submission  
JOURNAL Submitted (22-JAN-1998) Plant Sciences, University of Arizona,  
Tucson, AZ 85721, USA  
FEATURES  
source  
1..4818  
/organism="Arabidopsis thaliana"  
/strain="Ws-2"  
/db\_xref="taxon:3702"  
/chromosome="III"  
/map="8 cm from nga6"  
990..4398  
gene /note="DMF4"  
/gene="DMF4"  
/note="dwarf4"  
/allele="dwf4-1; dwf4-2; dwf4-3; dwf4-4"  
990..1055  
TATA\_signal  
/gene="DMF4"  
joinr<1133. .1353,1434. .1758,1844. .1996,2095. .2409,  
2562. .2654,2746. .2824,2931. .3040,3795. .4398)  
/gene="DMF4"  
Product="steroid 22-alpha-hydroxylase"  
join(1133. .1353,1434. .1758,1844. .1996,2095. .2409,  
2562. .2654,2746. .2824,2931. .3040,3795. .4040)  
/gene="DMF4"

/function="brassinosteroid biosynthetic enzyme"  
/note="member of the cytochrome P450 superfamily; CYP90B1"  
/codon\_start=1  
/product="steroid 22-alpha-hydroxylase"  
/protein\_id="AAC05093.1"  
/db\_xref="GI:2935342"  
/translating="MEFEHHTLPLPLPLPSLLPLPLILKRRNRKTRFNLPKGS  
GMPFGEITGYLKYPTATTLGDMQOAHKGYKIVRSNLGEPITVSADAGLNREILQ  
NEGRLEFCSYSPRISGIIIGKMSMLVLDGMDHMDRSISINFLSHARLRTILKDYERH  
TLFVLDMSMOQNSIFSAODEAKKFPFNMAKHMSMDRDEEETEOUKREYVPMKGVY  
APLNLPSTHRAKALDSRATILKFKIRKEERKLDIKEDQEEVEYKTEDKMSDSH  
VKRQRTDDDLGWLKHSNLSLSTOILDLILSLFAGHETSVALALAEFLQACPRAV  
BELREHDEISARAKKKESELSNMDYDKMDETCVIMETTLGNVVFRLHKAIDY  
RKGYDIPSGKVLPLVISAVIDNRSYDQPNLFNPMRMOQNNKASSGSGSFSTWGN  
NYMPFGGPRILCAGSELKLEVAVFILHVLVLFKNMELADEDQPPAFPFVDFNGLPIR  
VSRIL"

BASE COUNT 1566 a 727 c 888 g 1637 t  
ORIGIN

Query Match 35.4%; Score 1132; DB 12; Length 4818;  
Best Local Similarity 100.0%; Pred. No. 2.1e-137;  
Matches 1132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2071 gaattcaaatatcccatataatttttcgaatctcaaatatataattagcatt 2130  
Db 1 GAATTCAAATATCTCCATTAATATTTTTCAAATCTACAAATTAATTAATTAGTCAT 60  
Oy 2131 aacatgcatagaaagttccaaaaaaatttgtaacagaagaactccaaattttttt 2190  
Db 61 AACAAATGCATAGAAAGTCCAAAAAAATTTGTATAACAGAACTTCCAAATTTT 120  
Oy 2191 ttatagaaacagaataacagatagaacataatttggttgatgaaatgaaatga 2250  
Db 121 TTTATGGAACAAAGAAATTAACGATAGAAAACTATTTGTTGTGAATGAGATGAATA 180  
Oy 2251 tacattcaagcaattttaaanaattataagcctacagcgtcaagatgttctc 2310  
Db 181 TACATTAAAGCAATTTTAAAAATTTATTAAGCCTTATACCGCTCAAGATGTTATCTA 240  
Oy 2311 gtagtgtaattataatgcatgctgcatcagaattggagacaagaatgaagaat 2370  
Db 241 GTAGGTGATTAATTAATGATGATGCGATTCAGAAATGGGACAAACATGAAGCAAT 300  
Oy 2371 taataataactttaataataataataatttgatgaatgatttcttgacattgagg 2430  
Db 301 TAAATATTAACTTTTAAATAATAATAATTAAGTAATGTTGTTCTGACATTTGAGG 360  
Oy 2431 ggcataaagaagaacatgccaagaagtcacaggttgatgctcagttcgtaataatc 2490  
Db 361 GGCATAAAGAAAGCAATGCCAAAGCTTACGGGTTGACGTGCAGTCCGTAATATCT 420  
Oy 2491 aataactgctcttgacgcagcctgctgtagggatcctctgacaatttcaactgctc 2550  
Db 421 AATAACTGCTCTTTGACCGCACGCTGTGTAGGGGCTCTTGACATTTTCACTGTTCT 480  
Oy 2551 acccctactcgtgagcccaacccctttcccatatcctaaaggttaatttggaaatccaat 2610  
Db 481 ACCCTACTCTGTGAGCCACCTTTTCCATCTTAAGGATTAATTTTGGAAATCCCAAT 540  
Oy 2611 ttaacgcattgacgcagctacgcgaactccttggaattcctctgagacattatacaaat 2670  
Db 541 TTAAACCGATTGAGACCGTACCGGACTCTTGAGATTCTGCGAGCATTTATCAAAAAAT 600  
Oy 2671 tattagcaagaatgggttatttaattaaactcaaacatgatacagaataaattcat 2730  
Db 601 TATTACACGAATGGGTTTATTATTAAAACTCAACACTGATGATGAGATAAATTTCA 660  
Oy 2731 aaacactttacagatgattcgaacatcatatgaatttttttttttttttttttttt 2790  
Db 661 AAACACTTTTACATGATGATTCGATCACTATCTAATGACTTTTCTTCCACGAGT 720  
Oy 2791 ggaatgaagtatactactatagcagagaacaattgattatagataatccaatcc 2850

DB	721	GCATGAAGTTATCTACTATTAGCCAGACGAATGTATTATACATPATCCATTATTC	780
QY	2851	atgataattatgataataatgctgcttaactatttcaagcctgcagcttctgcgaact	2310
DB	781	ATGAAATTTATGATATTAATTAATGCTGTTAACTATTTCAGCTTCACCTTCTGCAACTT	840
QY	2911	ttgcttttaatttgaagcttcaataataaagatcttaaaagagagacgaagagacaca	2970
DB	841	TGCTTTTAAATTTAAGACTTTTATTAATAAAGATTTTAAAGAGGCGCTTAACGAGCACA	900
QY	2971	aaagaatgaacacgagagaaacaaagccatgaagctcaatggttgaagcttaata	3030
DB	901	AAAGTAATGAACACGCGAAGAAACAAAGCCATGAAGCTCATGTGTTAAGCTTATA	960
QY	3031	agaatatttctaatttaagtgagatgtaatacaattatcttgcacttttaaa	3090
DB	961	ACAAATATTTTATTAATTTTAATGACGATGTAACAAATATTAATTTTGCATCTTTTAAA	1020
QY	3091	acccctcttcaacaacagaagctccctcttccagtagaagtcgactcccaacttaag	3150
DB	1021	ACCCCTCTTTCACAAACAGAACCTCCCTTTTCAGTAGAAGTCGATTTCCCACTTTAAG	1080
QY	3151	acaaagccattagaagaagaagtgaagtagagagagagagaactctcc	3202
DB	1081	ACAAAGCCATTAGAAAGAGAAAGTGAGTGACAGAGAGAGAAACTGCTCC	1132
RESULT	3		
AC005504	104992 bp.	DNA	HTG 01-APR-1999
LOCUS	Plasmodium falciparum chromosome 12, *** SEQUENCING IN PROGRESS		
DEFINITION	***, 3 unordered pieces.		
AC005504			
AC005504.3	GI:4558584		
HTG:	HTGS, PHASE1.		
ORGANISM	malaria parasite P. falciparum.		
REFERENCE	Plasmodium falciparum		
AUTHORS	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
1 (bases 1 to 104992)			
Hyman, R.W., Fung, E.L., Qin, F., Tamaki, T., Kurd, O.B., Conway, A.B.			
and Davis, R.W.			
Plasmodium falciparum 3D7 chromosome 12			
2 (bases 1 to 104992)			
Hyman, R.W., Qin, F., Fung, E.L., Conway, A.B. and Davis, R.W.			
Direct Submission			
Submitted (21-AUG-1998) Stanford DNA Sequencing and Technology			
Center, Stanford University, 855 California Avenue, Palo Alto, CA			
94304, USA			
COMMENT			
On Apr 2, 1999 this sequence version replaced gi:4337172.			
* NOTE: This is a 'working draft' sequence. It currently			
* consists of 3 contigs. The true order of the pieces			
* is not known and their order in this sequence record is			
* arbitrary. Gaps between the contigs are represented as			
* runs of N, but the exact sizes of the gaps are unknown.			
* This record will be updated with the finished sequence			
* as soon as it is available and the accession number will			
* be preserved.			
1 58642: contig of 58642 bp in length			
* 58643 58842: gap of unknown length			
* 91012 91011: contig of 32159 bp in length			
* 91012 91211: gap of unknown length			
* 91212 104992: contig of 13781 bp in length.			
location/Qualifiers			
1. 104992			
/organism="Plasmodium falciparum"			
/db_xref="taxon:5833"			
/chromosome="12"			
BASE COUNT	44286 a	9326 c	9564 g 41411 t 405 others
ORIGIN			

Query Match	4.2%;	Score 133.2;	DB 60;	Length 104992;
Best Local Similarity	44.7%;	Pred. No. 7.6e-09;		
Matches 1041;	Conservative	0;	Mismatches 1248;	Indels 39;
				Gaps 12;

Qy	108	atataataatbaaaatcattcaattaaalatatcac-agtaataatlaattccttaacatg	166
Db	72449	ATTAAAGCTATTCAAAATATTTAAATATTAATAATCAATATTAATATTAATTTATTA	72508
Qy	167	tattgtataacaaaatcattcctcttggtatataatgaaaaatgatgagattggaaattac	226
Db	72509	ATTTAATAAATTAATATAAATATATTAATATTAATTAATTAATATCAAAATTAATCATTT	72568
Qy	227	aataataaaggaaataatcgatccattcgtgtgagat-tacaagttlaagttcttggtc	285
Db	72569	ATTAAATTTATTTTAAATAATATTAAATTTAAATATATATATATTAATTAATAATTAAGTT	72628
Qy	286	tctttgtctatatgtctatagagtaaaccaaaaagagatcgattgtaagtcgtacaacatct	345
Db	72629	AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	72688
Qy	346	tcgtctatgaccccccaaaaaaacaacaaca-aacaaaccccccccgatatacg	404
Db	72689	ACACAGTAAATTATATAAATAATGCAATTTATATTAATTAATTAATTAATTAATTAATTAATC	72748
Qy	405	cttttggtctcgagataggtcttaattgatacaataacatgacatcaattccttgattact	464
Db	72749	AATTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	72808
Qy	465	atbaagattctcttaccataaataattccgattcaatctcctctgtaataatlaataat	524
Db	72809	ATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	72868
Qy	525	acgagtcgtaataatccggttatacgatccactccaatcaatgataatgattcctgctgaatc	584
Db	72869	AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	72928
Qy	585	cagcaaatattataacaaagagatctgagaaaaaaccgaaaaataagaagaaggaagagtcg	644
Db	72929	TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	72988
Qy	645	tgaaccaatgagatctgtaataatlaaccaagaagaaatgaagatgacaaacaaagttctg	704
Db	72989	ATTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	73048
Qy	705	tggaaataatggtccctcgcaagctcttcctccacaatcaataatcgacctatttggattctc	764
Db	73049	ATTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	73108
Qy	765	tggatataatcgataaaatttgcgataacgatttgcgaaaaatcatttatttgcgtacgatac	824
Db	73106	ATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	73166
Qy	825	tcaataataatgctccaggtatttgcataactcttcgtttaaagcaatatttgccttctct	884
Db	73166	AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	73226
Qy	885	tttgtttgcttctctcttaactataatataatcgcgatataatgataacaaatgatatatacc	944
Db	73226	AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	73286
Qy	945	aaaacaatgtctcggagccaatttgcataaacttttctcaacaatlaaggagacactgga	1004
Db	73286	ATTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	73346
Qy	1005	cttcgacccctaaataacagatcttcaacggtctcctgattgagattctctgcataaagcata	1064
Db	73344	ATTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	73404
Qy	1065	aaggaccccgctcaagctatatacacaagttcaacaatgaatgacatgactgaaacccctta	1124
Db	73404	AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	73464
Qy	1125	gaaaatcttggaaataccggtctgtaatgaatatagatttagtggtcaacaatatgct	1184

[illegible]







Oy	1529	ttaacttgcttcctttaaagcaatccttiaacatttatgtgaagtggaaattaaga	1588
Db	73319	ATATATATTAAATTAATTAATAATAAATTTAAATTAATGATATATATATTAATTTTAT	73260
Oy	1589	gcgaacctgtaacaatlacaaatalctatacltagatacagtaigtgatctccaaataca	1648
Db	73259	TATATATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAC	73200
Oy	1649	taacttggatgctttaaacctaacttaactctgtttcttcctcgaagtaaaaatlaactcgag	1708
Db	73199	TATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	73140
Oy	1709	gtaaaaaaggcttgcctaatcttcogtagogaagaaggaataaccotaaigtacctaat	1768
Db	73139	ATATATTAATGTGTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	73080
Oy	1769	ttttgaaaaatgtaacccttlactcatagatatlaacogtaigtcttctgtgtgccaat	1828
Db	73079	ATTATTAATTAATTCCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	73020
Oy	1829	gacgagcccttaacaactgcygatagtccaattttcttcgcaaatcaatlaaggaatccaat	1888
Db	73019	TTTTAAATATGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	72965
Oy	1889	gctactatcaatagagaagaacagcgcgagiatlactaactttaacttaagacaacatttgg	1948
Db	72964	ATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	72905
Oy	1949	aaaaatgltataactttcttaacacatactlaaanaatgatagtccataatglatctccat	2008
Db	72904	ATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	72845
Oy	2009	gtctctaaatactttttttatactttagtgtcttaaatcacatagaaccaataaagtg	2068
Db	72844	TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	72786
Oy	2069	gtgaatccaatcatccatcataatacttttgaacatccaacatttaatacttaagta	2128
Db	72785	TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	72726
Oy	2129	ataacaatgcatagagaagtlccaaaaaaaatttcttaacgagaacatcccaaactttct	2188
Db	72725	ATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	72666
Oy	2189	ttttatggaaacagaagaataacagstlagaaacacttttctgttgysgaatgsgaagtaaa	2248
Db	72665	ATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	72611
Oy	2249	tatacatlaagcaaatlttlaaaaaatlcata	2280
Db	72610	TATATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	72579
RESULT	6		
AC004157/c			
LOCUS	AC004157	169546 bp	DNA
DEFINITION	Plasmodium falciparum chromosome 12 clone 3D7, *** SEQUENCING INCOMPLETE		
VERSION	AC004157		
KEYWORDS	AC004157		
SOURCE	HMG: HTGS-PHASE1.		
ORGANISM	malaria parasite P. falciparum.		
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.		
AUTHORS	1 (bases 1 to 169546)		
JOURNAL	Hyman,R.W., Fung,E.L., Qin,F., Rowley,D., Mao,J., Tamaki,T.,		
REFERENCE	Kurdi,O.B., Conway,A.B. and Davis,R.W.		
AUTHORS	Plasmodium falciparum 3D7 chromosome 12		
TITLE	Unpublished		
	2 (bases 1 to 169546)		
	Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.		
	Direct Submission		

JOURNAL

Submitted (19-FEB-1998) Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

COMMENT

On Aug 12, 2000 this sequence version replaced gi:8810447.

\* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as \* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 23466: contig of 23466 bp in length

\* 23467 23666: gap of unknown length

\* 23667 169546: contig of 145880 bp in length.

FEATURES

Source

1. 169546

/organism="Plasmodium falciparum"

/db\_xref="taxon:5833"

/chromosome="12"

/clone="PFYAC293"

/clone="3D7"

BASE COUNT 69871 a 15381 c 15705 g 68389 t 200 others

ORIGIN

Query Match 4.1%; Score 130.8; DB 60; Length 169546;

Best Local Similarity 44.3%; Pred. NO. 1.5e-08;

Matches 972; Conservative 0; Mismatches 1172; Indels 48; Gaps 9;

OY 106 agataactaataatataaaatcattcaataaataattcaagaftaataatcattacatc 165

Db 86323 AATATATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTT 86264

OY 166 gtagctatacaacaacatccatcttgtagtatacgagaatacgaggtcggaattca 225

Db 86263 ATTTAATTAAATTTAAATTTATTTAATTAAATTTAAATTTAAATTTAAATTTAATT 86204

OY 226 taataataaagaataatcgaatccatcttggtcgatcacagctcaagttcgtgct 285

Db 86203 TATTTATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAAT 86144

OY 286 tctttgtagtatagtatgaglaatacaaaaggtcttggttgaagtgtaacatact 345

Db 86143 AAAATTAATAATATATATTTATTTATTTATATATATTTAATTTAATTTAATTTATAT 86084

OY 346 tctgatacgaccccaaaaaaacaacaaacaaaccccccccgatagct 405

Db 86083 TTAATTAATTTAAATTAATTAATTAATTTAAATTTAATTTAATTTAATTTAATTTAAT 86025

OY 406 ttcttggtctcgatgaagttctatcttgatcacaattacatgcatctcttgatcacta 465

Db 86024 AATATATTTCAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAAT 85965

OY 466 tgaagatcttctcacaatlaaaatctcgaaatcactcctctgatacttaaatca--- 522

Db 85964 ATTAAATATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAAT 85905

OY 523 ----atacagagtgtgaataatccggttcatcgatcacatccaatcagatcatgactcttg 578

Db 85904 ATATATTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAAT 85845

OY 579 ctaaatccgacaattatcaacaagtgctgagaataaacggaataataagaagaaggaa 638

Db 85844 TTATTTACATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAAT 85785

OY 639 gagtctgacccatgagatcgtgataatcaatcaagaagataagagatgacaaccaa 698

Db 85784 TAAATATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAAT 85725

OY 699 aggtctgaggaataagtgccctcgcaagcttctctcacaatcaatcagaccctacttgg 758

Db 85724 ATATATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAAT 85665



QY 759 atttctcgatattcggttaaaattcgagatacagatgtgaaataatttattttag 818  
 Db 85664 TAATTTAAATATTAATTAAGA-----AAGATAATTTTATACCTTTA 85623  
 QY 819 cgaatcgaatctatgtccaggtatctgcataatctctgttaagcatttgc 878  
 Db 85622 TTTAATTTAAATATATAGTAATTAATTTTATGTTAT-TATTTAATTAATTTATTA 85564  
 QY 879 ttcttctgttctgtctcttaataatataatcgcgagataatgaataatgata 938  
 Db 85563 TTTTATTTTATTTATTTAATTAATTAATTTAATTAATTAATTTAATTTAATTA 85504  
 QY 939 tatcaaaaaaacattgtcggagccatttgaataacttttctcaaacatcaggac 998  
 Db 85503 ATATAAACAATTAATTAATTAATTAATTAATTAATTTTATTAATTAATTAATTA 85444  
 QY 999 actgagctcgacccttaataacgatttacaagcgcactagttgagattactagcata 1058  
 Db 85443 ATATTTTACATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 85384  
 QY 1059 agcataaagagccgltcaagctattatacaaaagtlacaactgaatagcttgaat 1118  
 Db 85383 AAAATAAATTAATTTATTTGTAATTAATTAATTAATTAATTAATTAATTAATTAAT 85324  
 QY 1119 ccttgaagaatttggaaattaccggttgttaagtaataatagattagtgtaacaaa 1178  
 Db 85323 TAAATTAATTAATTAATTAATTTATTTATTAATTAATTTTATTAATTAATTAAT 85264  
 QY 1179 ta-----tgttaacattagtggtgacacatacatcaatctctacagaaaaaaacact 1234  
 Db 85263 AATTATTTATTAATTAATTAATTAATTTTATTAATTTGTTAATTAATTTATTTATTT 85204  
 QY 1235 taagagaagtlacaatatacatatagtgatgctaaccttccagctacatactag 1294  
 Db 85203 TTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTT 85144  
 QY 1295 agcctaagaa-----tggtatgtagtgatgataaagaaatccacagcgtgtaac 1348  
 Db 85143 ATATATTAATTAATTAATTAATTTATTTATTAATTAATTAATTAATTAATTTA 85084  
 QY 1349 aactatggagccgatagttgcagtcacacatcatcattcgtgtgycacaataaa 1408  
 Db 85083 TTAATTTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 85024  
 QY 1409 aacaaaaaagaagaaagacgatttcttgatccatccaaatgacatcctaaatg 1468  
 Db 85023 ATTAAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTA 84964  
 QY 1469 catagatcttgggttaacggttgcgaagccctcaacagcgtgaacccctcgcaacta 1528  
 Db 84963 TTTAATTAATTTATTAATTTAATTTAATTTAATTAATTAATTAATTTAATTTAT 84904  
 QY 1529 ttaattgcttcttaaatgacatttaacataatttatttgaattgaattaaataga 1588  
 Db 84903 ATATATTTATTTATTAATTTATTAATTAATTTAATTTAATTTAATTTAATTTATTT 84844  
 QY 1589 ggcgaacttgtaacattacaattatatactagatagatgtaattccaataaca 1648  
 Db 84843 TAAATATTTATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 84784  
 QY 1649 tacttggagatttcaactaatctgttctctccctcaggtataaataatcatcagag 1708  
 Db 84783 TATAAATTTATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAT 84724  
 QY 1709 gtaaaaaaagtttcttcttctcgcagcagcgaaggaataaccacagacttaact 1768  
 Db 84723 ATATATTTATTTATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAT 84664  
 QY 1769 ttctgaaatgtaaccccttactcatagatlaattacgtaagtttcttgccacaat 1828  
 Db 84663 ATTAATTAATTTATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 84604

QY 1829 gacagccctacacactgtagtcaatatttctgcgaataatlaaaggaaatcaat 1888  
 Db 84603 TTTAATTAATTTATTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTT 84549  
 QY 1889 gctactatcaatagaagaacagcgtgagatcatcttcaatttaagaacaaatttgg 1948  
 Db 84548 ATTATTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTT 84489  
 QY 1949 aaaaatgtaataatcttcaacataatttaaaaatagatgcctataatgttccat 2008  
 Db 84488 ATTATTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTT 84429  
 QY 2009 gtcttcaaaatcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 2068  
 Db 84428 TTAATTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTT 84370  
 QY 2069 gtaacttcaaatctccatcaatatttcttgaactcaaaatlaataatlagta 2128  
 Db 84369 TAAATTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTT 84310  
 QY 2129 ataacaatgataagaagttccaaaataatttcttcaacgaacttccaaatttct 2188  
 Db 84309 ATTATTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTT 84250  
 QY 2189 ttcttgaagaagaataacagatagaaactatttcttcttgaactggaagtga 2248  
 Db 84249 ATTAAATTTATTTATTTATTTATTTATTTAATTAATTAATTAATTTAATTTA 84195  
 QY 2249 tatcataagcaaatcttcaaaatlatata 2280  
 Db 84194 TATATATTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 84163

RESULT 7  
 AEO01398 14867 bp DNA INV 06-NOV-1998  
 LOCUS Plasmodium falciparum chromosome 2, section 35 of the complete sequence.  
 DEFINITION  
 ACCESSION AEO01398 AEO01362  
 VERSION AEO01398.1 GI:3845197  
 KEYWORDS  
 SOURCE malaria parasite P. falciparum.  
 ORGANISM Plasmodium falciparum  
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 REFERENCE 1 (bases 1 to 14867)  
 Gardner,M.J., Jellellin,H., Carucci,D.J., Cummings,L.M., Aravind,L., Koonin,E.V., Shalimov,S., Mason,T., Yu,K., Fujii,C., Pederson,J., Shen,K., Jing,J., Aston,C., Lai,Z., Schwartz,D.C., Pertea,M., Salzberg,S., Zhou,L., Sutton,G.G., Clayton,R., White,O., Smith,H.O., Fraser,C.M., Hoffman,S.L. et al.  
 Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum  
 Science 282 (5391), 1126-1132 (1998)  
 JOURNAL  
 MEDLINE 99021743  
 ERRATUM: [[published erratum appears in Science 1998 Dec 4:282(5395):1827]]  
 REFERENCE 2 (bases 1 to 14867)  
 Gardner,M.J.  
 Direct Submission  
 Submitted (02-NOV-1998) The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20814, USA  
 JOURNAL  
 FEATURES  
 source  
 1..14867  
 /organism="Plasmodium falciparum"  
 /db\_xref="taxon:5833"  
 /chromosome="2"  
 complement(1570..2424)  
 /gene="PFB0490c"  
 complement(1570..2424)  
 /gene="PFB0490c"  
 /note="Predicted by Glimmer"  
 /codon\_start=1  
 /product="hypothetical protein"









ORGANISM Plasmodium falciparum  
REFERENCE Eukaryota: Alveolata: Apicomplexa: Haemosporidia: Plasmodium.  
AUTHORS 1 (bases 1 to 86827)  
Bowman, S., Lawson, D., Basham, D., Brown, D., Chillingworth, T.,  
Churcher, C. M., Craig, A., Davies, R. M., Devlin, K., Felwell, T.,  
Genette, S., Gilliam, R., Hamlin, N., Harris, D., Holtroyd, S.,  
Hornsby, T., Horrocks, P., Jagsels, K., Jassal, B., Kyes, S., McLean, J.,  
Moule, S., Muller, K., Murphy, L., Oliver, K., Quail, M. A.,  
Rajandream, M. A., Rutter, S., Skellon, J., Squares, R., Squares, S.,  
Sulston, J. E., Whitehead, S., Woodward, J. R., Newbold, C. and  
Barrell, B. G.  
TITLE The complete nucleotide sequence of chromosome 3 of Plasmodium  
falciparum  
JOURNAL Nature 400 (6744), 532-538 (1999)  
MEDLINE 99376085  
REFERENCE 2 (bases 1 to 86827)  
AUTHORS Bowman, S., Skellon, J., Churcher, C., Lawson, D., Quail, M. and  
Barrell, B.  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 86827)  
AUTHORS Lawson, D., Bowman, S. and Barrell, B.  
TITLE Direct Submission  
JOURNAL Submitted (17-DEC-1998) P. falciparum Genome Sequencing Consortium,  
The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge  
CB10 1SA, UK  
COMMENT On May 5, 2000 this sequence version replaced gi:4493931.  
For more information about this sequence or the Malaria Project,  
see [http://www.sanger.ac.uk/Projects/P\\_falciparum](http://www.sanger.ac.uk/Projects/P_falciparum).  
FEATURES  
source location/Qualifiers  
1..86827  
/organism="Plasmodium falciparum"  
/strain="3D7"  
/db\_xref="taxon:5833"  
/chromosome="3"  
/clone="MAL3P5"  
gene join(322..603,826..1023,1197..1301,1458..2942)  
/gene="PFC0575w", MAL3P5.1"  
join(322..603,826..1023,1197..1301,1458..2942)  
CDS /gene="PFC0575w", MAL3P5.1"  
/note="PFC0575w (MAL3P5.1), Hypothetical protein, len: 689  
aa, possible signal sequence, revised: added new exon 2"  
/codon\_start=1  
/product="Hypothetical protein, PFC0575w"  
/protein\_id="CAB38969.2"  
/db\_xref="GI:7711065"  
/db\_xref="SPTREMBL:O97258"  
/translation="MYLKNVYIYISCFILDFCLPSFHLKMKYKNNNNKSVTFEL  
KSPQYLRKFRKRSKKNVSEFKKOKKPLFLENLKGFSFLGFWNODKYIDVIS  
NINNLTRIKOYTHKKKSNSEFTKENTKILHCVSFKIDFKIINNLSYIKRFOMSNIT  
VHSILNOISEKKEKDAENYLAHLFLKIDENITLFSMMHIMDFESKQKVCIEIRD  
IKSKRRKKNLSIYNLFCILYETVCMCLIKYISHLGIEFFFEFCFCYCNILIER  
IYEEVGDILIRKTERVNLVCEKKIKRPHMDAIIKKMINKMDIYFNHYDELICR  
FTMKLNIERNKKNIIIRSYNDINNDISIDKMYMNNPDIYVNIINNISLDEKJKEQEPN  
DDENLKELDYEQFLFNDNIIKTYIEDOPLVYNDSNINNDNNNNINTKNNKHKIK  
DYVNDDDDYDEKEDLVYIKNIDYIYKNTIGMKSLEEKNOFIEQADIEFQNFLS  
NVLNDQHGKRVSNDEKSTEHKIKNKTINNGYETLIONMENNFIKKNIDNINSND  
ISNNENIIRIKLKNOSDEINLSDLYERLTKRVLWYIKQIKELKRYQYDIINEQ  
YPIIKNEKTVLDLNTGYKIYMSPDVNSLPEFKIKIDISIPREKKNQNMENQKSKNY  
IMERNVM"  
misc\_feature 604..609  
/note="predicted splice donor sequence for exon 1 of  
PFC0575w (revised)"  
misc\_feature 816..825  
/note="predicted splice acceptor sequence for exon 2  
(revised of PFC0575w)"  
misc\_feature 1024..1029  
/note="predicted splice donor sequence for exon 2 of  
PFC0575w (revised)"  
gene /gene="PFC0580c", MAL3P5.2"  
complement(3354..6644)  
CDS /gene="PFC0580c", MAL3P5.2"  
/note="PFC0580c (MAL3P5.2), Hypothetical protein, len:

1097 aa, possible signal sequence, predicted using  
hexagon.  
/codon\_start=1  
/product="Hypothetical protein, PFC0580c"  
/protein\_id="CAB38968.1"  
/db\_xref="GI:4493932"  
/db\_xref="SPTREMBL:O97257"  
/translation="MLGLKRRNFYLLVSPSLFAYFLKRRHKNENNYETLINNDIE  
KIKRIRKNSYIPFLFLNLYDSYIKNLIRMLYEFERRRKRKKEYYIYTNVVK  
KRREAIKRYNLESDQNLNFKRYIYEIVLSEIKGILSPHLYLTKISHCVAIYR  
SLYYNKLNDKNNHLNLEKKLKYFPOINNEHQOAPTNTTHNNNNNNKKPDIINSHC  
KNYNISSTYNNNEKENINIDRYNHNFTESISYKDEKCOHITLMMYLLNQTY  
DNICRCLNTNINIIYINFMINILKYICQKMEIILLYNHIIDKKKTIQKNQNTS  
LEKYIYSEFFKKEENHIYDFEDQMHLHKKENDKYVNSNETNHNIIKYISDN  
FYDHINSSNRCSPFKNLKQOQDQDQTKHIIKGKEKYPNKKNSHKKNNQNTIEK  
DQKDIILKYIYKGNKIDDIOLINELVYMYMLLFPSCSLISIKKHILKEKME  
FSDKNTIYLSADPMNLRNRIILKRSKNEBEREINFSFASPFILSKNIIYFDEICG  
RSRDNTIYNAVYDETKNTYNNNNNDNDNDCSNNDHICSNNDHICSNNDHICSN  
NNNNICSNNNNNICSNNNNNICSNNNNNICSNKMLDEQCQNFNDYTRKREKRI  
YELAKIYNNIYEDYLGKREKHONEDNTINLYIKKCPMAYFLKINIKNDTSIEH  
NNNIYNGDINKNNIIFKKKYNLPESIIISYIKDIYENYKRLRYIYDNLIKKCEK  
YELKMEHINRKLKMKRAPHYIYNFPOFIINNYIHKNNIHKIHIHKKCKDE  
IDYKFDLYICMINNINISYIHKVDHNECVRIFFAYNKKILYETNINLEKENIT  
YKMKIKYLYLNNNISMDLYPNISYKRIYNQKKYKRNKFSHETYSLKNDIHLFL  
LYTORIQNCDFISYIYKKNYNEKNPFLNLYELAHYVSEKKKKKFFSSTSSP  
YSYDMVNSFTFFSFFSLYLLFLFYHPDMYASYIFFKTLTYSGLPTYYSYNNI  
MYVCGPKR"  
gene join(7669..7717,7800..7829,7912..7940,8064..8106,  
8302..8368,8568..8641,8713..8812,8924..9006,9122..9258,  
9369..9505,9613..9838)  
CDS /gene="PFC0581w"  
join(7669..7717,7800..7829,7912..7940,8064..8106,  
8302..8368,8568..8641,8713..8812,8924..9006,9122..9258,  
9369..9505,9613..9838)  
/gene="PFC0581w"  
/note="PFC0581w, hypothetical protein, len: 324 aa,  
revised: new gene prediction, splicing prediction very  
tentative"  
/codon\_start=1  
/product="Hypothetical protein, PFC0581w"  
/protein\_id="CAB90285.1"  
/db\_xref="GI:7711066"  
/translation="MKKKKHFFSIFKPVNDKNNKYLCSKDYIRIINMYIGLHIFRL  
OKFIHYPANQIASSPFESHKNRYVEDIIDGKEKMFNTDILIESIKQNDPFS  
NDIKTYVAMKNNKMDQLTYSNKKIDTYVANSQWQASPYFLINIKFSRMSQPA  
LKVDEELVSKKNNPSALSDNSQYKTYIVTLTDNIESTKYNFASQYGVY  
TLKREKRIINRLLSKERYPMQWMDMKERKHITTFYVITNLFELISLHRHDSV  
QNFLEKKNNSDKLQDDIDEDERYFDEILIREAKKSEYDKODEL"  
misc\_feature 7790..7799  
/note="potential splice acceptor sequence for exon 2 of  
PFC0581w"  
misc\_feature 7830..7835  
/note="potential splice donor sequence for exon 2 of  
PFC0581w"  
misc\_feature 7901..7911  
/note="potential splice acceptor sequence for exon 3 of  
PFC0581w"  
misc\_feature 7941..7947  
/note="potential splice donor sequence for exon 3 of  
PFC0581w"  
misc\_feature 8055..8063  
/note="potential splice acceptor sequence for exon 4 of  
PFC0581w"  
misc\_feature 8107..8112  
/note="potential splice donor sequence for exon 4 of  
PFC0581w"  
misc\_feature 8293..8301  
/note="potential splice acceptor sequence for exon 5 of  
PFC0581w"  
misc\_feature 8369..8374  
/note="potential splice donor sequence for exon 5 of  
PFC0581w"  
misc\_feature 8559..8567



Oy	1544	taagcatttcaacattcttcgttaagtcggaattcaagaacggaactcgtaac	1603
Db	39930	TAAATTATATATCAATTTAAATGTTATTA-TAATATTATTAATAATTTAAATTTAT	39877
Oy	1604	tacaattattacatagactagtaatgtaattccaacatacacttggatgtta	1663
Db	39871	TAAATTATATTTTAAATATATTATTTAAATTTATTTATTAATATATTAATTAAGTGA	39811
Oy	1664	aactcaactcgttctctccctcgcgtataaactcaatcgcgggtaaaaagcttcg	1723
Db	39811	ATAAAAATATAAATATATTTATTT-----TATATATATATATTTATTTATTAATACT	39766
Oy	1724	tctatttcgcgacgcgaaggataaacccatgcacttaatttcgaaagtaac	1783
Db	39739	GCTTATTTAATTCATTATTAATATATATATATTAATTTGCAATTAATTAATTTATTTAT	39700
Oy	1784	cccttactcatagatlaaatccgtaigtcttgcgtcaataagacagccctcaac	1843
Db	39639	TATAAATTTATTTATTAATTAACATTTTATTTAT-TTAATTTATTTATTTATTTAATTA	39644
Oy	1844	tgtgtagcaactttcttcgcaaatataaattggaattcgaactcactcactaga	1903
Db	39640	TATTATATATATATTTATTTGCAATTTATTTATTTATTTATTTATTTATTTAATTAAT	39588
Oy	1904	agaaacagcgcgaglatcacatttcaattcaagaacaaaatttcgaaaaagtacaa	1963
Db	39580	ATATTTTATTTATTTATTTTAAATTAATTTATTTATTTATTTATTTATTTAATTAATTA	39522
Oy	1964	tctaacatattatcaaaatagatgcccataatgatttccctagtccttaaaatct	2023
Db	39520	TTAATTTATTTATTTAATTAATTAATTTATTTATTTAATTAATTAATTAATTAATTT	39461
Oy	2024	tctttacatttagtataaatacacatlaagaccaaataagtcggtgaatccaatc	2083
Db	39460	TATTTTATTTAATTAATTAATTAATTAATTTATTTATTTTAAATTAATTAATTA	39400
Oy	2084	tcacattaatatttttggaaatccacaattataatatttgatcaataaactgcata	2143
Db	39400	TTAATTTATTTATTTAATTAATTAATTAATTAATTTATTTATTTTAAATTAATTA	39341
Oy	2144	aagttccaaaaaaatttgcgttaacggaactccaattcttttctlaagaaag	2203
Db	39340	TATATATTAATTTATTTATTTTAAATAAATAATATATTAATTTATTTATTTAAT	39281
Oy	2204	aaatc---acagataagaaactacttgcgtcggaatgggaagtagtaatacatgaac	2260
Db	39280	AAATTTATATATATTTATTTATTTTAAATAAATAATTAATTAATTTATTTATTTT	39221
Oy	2261	aaatttcaaaaattataaagccatacagcgctcaagaatgtaactgaagtagta	2320
Db	39220	AATTAATATATATTTTATTTTATTAATTTAATTTAATAATTTAATTTATTTTGTAT	39161
Oy	2321	ttaa--taatgcatagtgcgaltcgaacttgggacacaatgaaacggaatlaaata	2378
Db	39160	TTTATTTATTTTAAATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTAAT	39101
Oy	2379	taacttcaaaataaataaaatttgagcaaatgigtcttcgacat	2425
Db	39100	TTAATATTAATTAATTAATTAATTAATTTATTTTATTTATTTATTTT	39054

RESULT	10			
DMU11584				
LOCUS	DMU11584	4601 bp	DNA	INV
DEFINITION	Drosophila melanogaster Oregon-R mitochondrial A+T region.			
ACCESSION	U11584			
VERSION	U11584.1	GI:508826		
KEYWORDS	mitochondrial DNA, A+T region, tandem repeats.			
SOURCE	fruit fly.			
ORGANISM	Mitochondrion Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;			

REFERENCE	AUTHORS	TITLE
1 (bases 1 to 4601)	Muscomorphia: Ephydroidea; Drosophilidae: Drosophila.	
2 (bases 1 to 4601)	Lewis, D. L., Farr, C. L., Farguhar, A. L. and Kaguni, L. S.	Sequence, Organization and Evolution of the A+T Region of Drosophila melanogaster Mitochondrial DNA
Mol. Biol. Evol.	11, 523-538 (1994)	
JOURNAL		
MEDLINE	94285822	
REFERENCE	2 (bases 1 to 4601)	
AUTHORS	Kaguni, L. S.	
TITLE	Direct Submission	
JOURNAL	Submitted (28-JUN-1994)	
FEATURES	Laurie S. Kaguni Ph.D, Dept. of Biochemistry, Michigan State University, East Lansing, MI, 48824-1318, USA	
	Location/Qualifiers	

[illegible]



OY	1311	atgagatgctgaataaagaattccacacgcggtgaataatc---atggacgcgtatgt	1367
Db	3521	AAATTTATTTAAATTAATTTAAATTAATTTTAAAAATTTCTTAATGATATTTTAT	3580
OY	1368	acgaacactgcgaatcatcctcttgctgcacacataaaacacaaagaanaaa	1427
Db	3581	AAAAAATTTTATATTAATTAATAATCAGCTTTTAAAAAATTAACAAAATTTTAAATA	3640
OY	1428	gaaaacgattcttccttgatccatcgaatgactaaatgcatagatcctctgggttac	1487
Db	3641	ATTAATTTTAAATGAATTAATTAATTTATTTATTTTCAATTTTAAAAAATTTTAA	3700
OY	1488	agttcgaaagccctcacaacgcgtgaacacatcgaaatlaaattgcttcctc-ttaa	1546
Db	3701	AAAAATATATTTTTTTAAAAAATACTATATCTAATTTTAAATTAATATGATATTTA	3760
OY	1547	tgcatccttaacacatattatcttgatcttggaatttaagaacgaacttgtaacatcac	1606
Db	3761	TATATATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	3820
OY	1607	aatatctatctagatcactcgtatgcatatccaaatacactcttgagatgtttaaac	1666
Db	3821	AAATTTAAAAATGATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	3880
OY	1667	ctaaccttgcttcctcctacgataaataatcatcagcgataaaagaattgctct	1726
Db	3881	TTTATTTTATTAATTAATTAATTAATGATTTTATTAATTAATTAATTAATTAATTA	3940
OY	1727	tatttcgcgaatgaagaagataaacctaactgaattacttattcttgaaatgtaacct	1786
Db	3941	ATTATTCGAAATTTATTAATTAATTAATTAAT-ATTAATGAGAAAATTAATTAATTA	3999
OY	1787	cttactcatgattaaatcaccgatgttcttgctgcataagacagcctctacaacgt	1846
Db	4000	TTTATATTAATTTTAAAAAATTTCTTAATGATTAATTTTAAAAAATTTATAT	4059
OY	1847	gatagtcgaattcttcgcgaacatlaaatlaggaatcgaatgcatacgaataaga	1906
Db	4060	AATTAATAATCATTTTTTTAAAAATTAACAAAATTTTAAATTAATTAATTTTAAATG	4119
OY	1907	aacgcgtagatlatcacattlaaattlaagaacaaatcttgaaaatgltataactct	1966
Db	4120	AAATATATTAATTTATTTTTCATTTTAAAAAATTTTAAAAAATTAATTTT	4179
OY	1967	aacaaatattataaatacagatgcctataatgattctctcctatgctcttaaatattct	2026
Db	4180	TTTTAAAAAATACTATATCTAATTAATTAATTAATTAATTAATTAATTAATTAAT	4239
OY	2027	cttatattatgataaatacatatagaaccaaataatgcttgatcaatcaatactcc	2086
Db	4240	TTAATATTTTATTAATATCTAATTAATTTAAATTAATTAATTAATTAATTAATTAAG	4299
OY	2087	altaatatcttggaaactcacaatataatataattatgtaacaaatgcatagaag	2146
Db	4300	ATATATATTAATTAATTAATTAATTTCCATATTAATTTATTTAATTAATTAATTA	4359
OY	2147	cttccaaaaaaatcttgcttaacggaactccaaattctttcttctatgagacaaga	2206
Db	4360	TAAATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	4419
OY	2207	taacagatagaanaactcttgcttgtaggaatggaatgataatacatagaacaattc	2266
Db	4420	TATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	4479
OY	2267	taaaaaatlatataagcctacacgcgcgcgaaglatgltactctagtaggtlaata	2326
Db	4480	AAAAAATTTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	4539
OY	2327	atgcgtggtgcgtcgaacttgtagcacaacaaatgaacaggaattcaaatctta	2386
Db	4540	TTTTTTAAAAAATAAGTAATTAATTAATAAATAAATAAATAAATAAATGAAATTA	4599

```

RESULT 11
LOCUS      DMU37541      19517 bp      DNA      circular      INV      02-MAR-2001
DEFINITION Drosophila melanogaster complete mitochondrial genome.
ACCESSION  U37541
VERSION     03/541.1   GI:1166529
KEYWORDS
SOURCE
ORGANISM   fruit fly.
            Mitochondrion Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 12511 to 12682)
            Clary,D.O., Goddard,J.M., Martin,S.C., Fauron,C.M. and
            Wolstenholme,D.R.
            Drosophila mitochondrial DNA: a novel gene order
            Nucleic Acids Res. 10 (21), 6619-6637 (1982)
TITLE
JOURNAL    2 (bases 5269 to 5695)
MEDLINE    83090428
REFERENCE  3 (bases 404 to 5272)
AUTHORS    de Bruijn,M.H.
            Drosophila melanogaster mitochondrial DNA, a novel organization and
            genetic code
            Nature 304 (5923), 234-241 (1983)
JOURNAL    4 (bases 804 to 1778)
MEDLINE    83245048
REFERENCE  5 (bases 14215 to 14512)
AUTHORS    Ballard,J.W., Olsen,G.J., Faltch,D.P., Odgers,W.A., Rowell,D.M. and
            Atkinson,P.W.
            Evidence from 12S ribosomal RNA sequences that onychophorans are
            modified arthropods
            Science 258 (5086), 1345-1348 (1992)
JOURNAL    6 (bases 441 to 2967)
MEDLINE    88212147
REFERENCE  7 (bases 14215 to 14512)
AUTHORS    Saita,Y. and Takahata,N.
            Evolution of Drosophila mitochondrial DNA and the history of the
            melanogaster subgroup
            Proc. Natl. Acad. Sci. U.S.A. 87 (24), 9558-9562 (1990)
JOURNAL    91088557
MEDLINE
REFERENCE  9 (bases 1 to 408; 13319 to 19517)
AUTHORS    Lewis,D.L., Farr,C.L., Farquhar,A.L. and Kaguni,L.S.
            Sequence, organization, and evolution of the A+T region of
            Drosophila melanogaster mitochondrial DNA
            Mol. Biol. Evol. 11 (3), 523-538 (1994)
JOURNAL    94285822
MEDLINE
REFERENCE  9 (bases 1 to 408; 13319 to 19517)
AUTHORS    Lewis,D.L., Farr,C.L., Farquhar,A.L. and Kaguni,L.S.
            Drosophila melanogaster mitochondrial DNA: completion of the
            nucleotide sequence and evolutionary comparisons
            Insect Mol. Biol. 4 (4), 263-278 (1995)
JOURNAL    96423163
MEDLINE
REFERENCE  10 (bases 1 to 19517)
AUTHORS    Lewis,D.L., Farr,C.L. and Kaguni,L.S.
            Direct Submission
            Submitted (03-OCT-1995) Laurie S. Kaguni, Biochemistry Department,
            Michigan State University, East Lansing, MI 48824-1319, USA

```

```

FEATURES
source
location/Qualifiers
1..19517
/organism="Drosophila melanogaster"
/organelle="mitochondrion"
/db_xref="taxon:7227"
/note="derived from new and previously submitted
sequences; sequence is a composite containing sequences
obtained from different Drosophila melanogaster strains"
1..65
/product="tRNA-Tle"
/complement(97..165)
/product="tRNA-Gln"
171..239
/product="tRNA-Phe"
240..1265
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 2"
/protein_id="AAC4781.1"
/db_xref="GI:1166530"
/translation="MFNNSKILFTITIGTLITVTSNSWLGAMGLETINLSPIPL
LSDNNNMSTASLKYRTQVLASTVILFSSILLMKNNNNNESTSMINSALL
LKSQAPFHEWEPNNMEGLTWNALMTQKIAPLMLISTLNTKILLSVILSVII
GAIGLNOTSLRKLMAFSSINHLGWLMSLMSISEIWLILFEFYSFSEVLTPEMFI
KLFLNQLFSWFVNSKILKFLFMNFIISLGLPPLGLPLPMVLVIOQLTLCNOYFMLT
IMMSTLITLFEYLRICYSADPMNFEENNIMKMNINSINMYIMTFISIFGLFI
SLFVFMF"
1264..1329
/product="tRNA-Tyr"
/complement(1322..1383)
/product="tRNA-Cys"
/complement(1403..1468)
/product="tRNA-Tyr"
join(1470..1472,1474..3009)
/note="mechanism underlying reading frame shift after
first codon uncertain"
/codon_start=1
/transl_table=5
/product="cytochrome c oxidase subunit I"
/protein_id="AAC47812.2"
/db_xref="GI:7412849"
/translation="MSRWLSTNHNKDIGTLPTFGAGAGVGSLSLILAEGLHPC
ALIGDQIYNIVTAHAETIMIEFVMPIMIGFGNWLVPMLGADMAFPRNNNSFW
LLPAPLSLLVSNVENGACGTGVYPLPSAGIAGGASVDLAIFSLHLAGISSILGA
VNEITTVNMRSTGISLDRMPLEWVSIVTILLLSLPVLGATMLLTDRNMTSE
FDPGGGDPIILYOHLEWFEFGHEVYILILPGEGMSHIIISQSGKKEFGSLGMYAM
LAIGLIGTVAHHMTYGMVDPRAYTSMTIAYPTGIRKITSMTLTLCTOISYS
PAIIMALGFVFLFTVGLTGVLANSSVDIILHDITVVAHFHYLVSGAVFALMAGF
IHWPPLFTGLTLNKKMLKSHPIIMEIGVNLFFPDHFLGMPRRSYDPAATYTN
IVSTIGSTISLGLIFFEFTIMESLVSORQVIYPIQLNSSIEMVQNPAPHSYSELP
LLTN"
3012..3077
/product="tRNA-Leu"
3083..3767
/note="TAA stop codon is completed by the addition of 3' A
residues to the mRNA"
/codon_start=1
/transl_except=(pos:3767,aa:TERM)
/transl_table=5
/product="cytochrome c oxidase subunit II"
/protein_id="AAC47813.1"
/db_xref="GI:1166532"
/translation="MSTWANGLDOSASPLMEOLIFFPHDALLILVMTTVAGVLMFM
LFFNNVVRRLHGOILIEWITLIPALIFALPSRLYLDEINPSVLTSGICH
OWMSYNSDPNNIEFDSYMTPTNELMDGRLLDQNRVYLPMNSQIRILVTADYI
HSWTVPALGVAVDSTPGRLNOTNEFINRPLFLYGQCSICGANSFMPIVIESPVN
FIKISSNNS"
3768..3838
/product="tRNA-Lys"
3840..3906
/product="tRNA-Asp"
3907..4068
/codon_start=1

```

```
/transl_table=5
/product="ATPase 8"
/protein_id="AAC47814.1"
/db_xref="GI:1166533"
/translation="MPOMAPISWLLIFIFSITPLFCSINYISYMPSPKSNELKNI
NLSNMNMW"
4062..4736
/codon_start=1
/transl_table=5
/product="ATPase 6"
/protein_id="AAC47815.1"
/db_xref="GI:1166534"
/translation="MMTNLFVFDPLAIFNFSNLNWLSTFLGLMIPSYIWMPSRYNI
MNSILILHKEFKTLPGSCHNGSTFIIFSLILPNNMGLEPYIFETSHLTL
LSLALPLMELGASWPMGIISFNFOIPLNTATLLASGVTVAHSHLMENFSOT
TQGLFETVLIGYFTIIQAYEIEAPFTIADISYSGTFPMAGFHIVLIGTTFLLV
CLRLHNHFSKNHFGFEAAMWIMFVDMVWLFITITIMWG"
5543..5607
/product="CRNA-Gly"
5608..5961
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 3"
/protein_id="AAC47817.1"
/db_xref="GI:1166536"
/translation="MFSIFIALILLITTYIMFASILSKKALIDREKSPPECCPD
PKSSRLPELRFLLITIIPLIFDVEIALILPMIIMKYSNMIMWTISILFILLLI
GLYHMOGMNMSN"
5981..6045
/product="CRNA-Ala"
6055..6118
/product="CRNA-Arg"
6119..6183
/product="CRNA-Asn"
6184..6251
/product="CRNA-Ser"
6252..6318
/product="CRNA-Glu"
6319..6401
/product="CRNA-Phe"
6402..6401
/translation="TAA stop codon is completed by the addition of 3' A
residue to the mRNA"
/codon_start=1
/transl_except=(pos:complement(6401..6402),aa:TERM)

Query Match      3.9%: Score 124.4, DB 6; Length 19517;
Best Local Similarity 43.9%: Pred. No. 1.3e-07;
Matches 1027; Conservative 0; Mismatches 1276; Indels 37; Gaps 10;
```

```

OY 246 gattccattggttgatcacacagtgatttggcttcttctgtatataatgatac 305
DB 17372 TTTATTACCAATAATTAAATTAATAAATAAATTAATTTTCTTTTCTTTTATA 17431
OY 306 agtataacaaaagagatgattgaagtgtaacatattcgatagcccccaaaaa 365
DB 17432 ATTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 17491
OY 366 aaaaaaaacaaacaaacaaaccccccccgatagatttggcttgataggtt 425
DB 17492 ATATATATGAAAAATTA-----AATTAATTAATTAATTAATTAATTAATTAAT 17536
OY 426 tattatcataataacatgacatcattcttgattcattcaagaatttcttaccatt 485
DB 17537 TTTTAAATAATTTCTTAATAGTATTAATTTTATAAATAAATTTATTAATTAATTAATGCT 17596
OY 486 aaaaattcgatcatatctcttgatttaataataacagagtgtaataatcggtta 545
DB 17597 TTTTAAAAAATTAACAAAAAATTTTAAATAAATTTTAAATAATTAATTAATTAATTAAT 17656
OY 546 tggatcaccccaatcatgattatgattcttgcttaatccagaattattacaagaat 605
DB 17657 TATATTTTCATTTTAAAAAATAATTTTAAAAAATAATTTTAAAAAATAA 17716
OY 606 attgaagaaaaacccgaataaagaagaaaggaagtagtgaccatggagatggaat 665
DB 17717 CTAATACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 17776
OY 666 aattacaagaagaataagaagatgacaacaaaggcttggaataatggtccctgccc-- 723
DB 17777 ATATATCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 17836
OY 724 agcttctctcaacaatcatatcgaccattcggttcttggaataatcgataatc 783
DB 17837 AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 17896
OY 784 ggcgaacagattgga--aaattatttcttgagcgaatcgaatcgaatg--tcca 840
DB 17897 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 17956
OY 841 ggtattgataatctctctgttaagaacatatttcttcttcttcttcttctct 900
DB 17957 TATATATGAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 18016
OY 901 taactataattatcgcgatata-----gataacaatgatatcacaaacattgt 955
DB 18017 AATGATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 18076
OY 956 ctgggacacatttgaataacatttctcaaacattcggagacacgagctgacccctta 1015
DB 18077 AAAAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 18136
OY 1016 aaatcagatttcaacgctcatcagtttgatgattactagacataagaagaccggt 1075
DB 18137 TTTAAAAAATTTTAAAAAATAATTTTAAAAAATAATTTTAAAAAATAATTAATTAATTAATTA 18196
OY 1076 ca-----agctatttacaagaattacaacagtaaatgctgaataatccttaagaat 1130
DB 18197 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 18256
OY 1131 ttggaattacccggttgattgtaaatatagatttggtgtaacaataatgtaatac 1190
DB 18257 AATTAATAAATTTTAAATTTTAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 18316
OY 1191 ttagtgcaacatataatattccttacaagaataacaaacttaagaagaatlaaacat 1250
DB 18317 ATTTATTTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 18376
OY 1251 atccatatatgggtatgctatcattccttcaagttatgctatcagtaagaacaaatagtt 1310
DB 18377 ATTTAATAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 18436
OY 1311 atgtgatgctgataaataatcacacgctgtaataat--atgggaaccgatagtt 1367
```

```

Db 18437 AATATATTTAAATTAATTAATTAATTTTAAAAATTTCTTAATGATATTATTAT 18496
Oy 1368 aagatcagcgaataatcctcttggtgcaacataaacaagaagaaaaa 1427
Db 18497 AAAAAATATTTATATATATATATATATATATATATATATATATATATAT 18556
Oy 1428 gaaaacgattctcttgatccatcgaatgatacgaatgatacgaatgatac 1487
Db 18557 ATAAATTTATATATATATATATATATATATATATATATATATATATATAT 18616
Oy 1488 agttcgaagtcctcacaagcgtgaaccatcgaactgaactgaactgaact 1546
Db 18617 AAAAAATATTTTATATATATATATATATATATATATATATATATATATAT 18676
Oy 1547 tgcattcttaacatattcttgatggaattgaatgaatgaatgaatgaatga 1606
Db 18677 TATATATATATATATATATATATATATATATATATATATATATATATAT 18736
Oy 1607 aatattatattagatagatagatagatagatagatagatagatagatag 1666
Db 18737 AATTTAAAAATGATATATATATATATATATATATATATATATATATATAT 18796
Oy 1667 tcaacttgctctcctcctcgaatgaatgaatgaatgaatgaatgaatga 1726
Db 18797 TTTATATATATATATATATATATATATATATATATATATATATATATAT 18856
Oy 1727 tcttctcgaatgcaatgaatgaatgaatgaatgaatgaatgaatgaatga 1786
Db 18857 ATTATTTGATATATATATATATATATATATATATATATATATATATATAT 18915
Oy 1787 ttactacatagatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 1846
Db 18916 TTTATATATATATATATATATATATATATATATATATATATATATATAT 18975
Oy 1847 gatagtcgaatcttctcgaatgaatgaatgaatgaatgaatgaatgaatga 1906
Db 18976 AATTAATATATATATATATATATATATATATATATATATATATATATAT 19035
Oy 1907 aacagctggaatcttctcgaatgaatgaatgaatgaatgaatgaatgaatga 1966
Db 19036 AATATATATATATATATATATATATATATATATATATATATATATATAT 19095
Oy 1967 aacatattatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 2026
Db 19096 TTTTAAAAAACTATATATATATATATATATATATATATATATATATATAT 19155
Oy 2027 ttatattagatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 2086
Db 19156 TTAATATATATATATATATATATATATATATATATATATATATATATAT 19215
Oy 2087 attaatcttctcgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 2146
Db 19216 ATATATATATATATATATATATATATATATATATATATATATATATATAT 19275
Oy 2147 ttcgaataaataatcttctcgaatgaatgaatgaatgaatgaatgaatgaatga 2206
Db 19276 TAAATATATATATATATATATATATATATATATATATATATATATATAT 19335
Oy 2207 taacgatgaataatcttctcgaatgaatgaatgaatgaatgaatgaatgaatga 2266
Db 19336 TATATATATATATATATATATATATATATATATATATATATATATATAT 19395
Oy 2267 taagaataatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 2326
Db 19396 AAAAAATTTCTTAATGATATATATATATATATATATATATATATATATATAT 19455
Oy 2327 atgcataatgcatgaatggaatggaatggaatggaatggaatggaatggaatgga 2386
Db 19456 TTTTAAAAAAATATATATATATATATATATATATATATATATATATATAT 19515

```

RESULT 12

```

PFMALIP3/C
LOCUS      67970 bp      DNA      INV      15-DEC-1999
DEFINITION Plasmodium falciparum MALIP3, complete sequence.
ACCESSION  AL031746
VERSION     AL031746.9  GI:6594243
KEYWORDS   HTG.
SOURCE      malaria parasite P. falciparum.
ORGANISM   Plasmodium falciparum.
REFERENCE  1 (bases 1 to 67970)
AUTHORS    Bowman, S., Churcher, C., Harris, B., Harris, D., Lawson, D., Quail, M.
            and Barrell, B.
TITLE       Direct Submission
JOURNAL     Submitted (24-SEP-1998) P. falciparum Genome Sequencing Consortium,
            The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA, UK
COMMENT     On Dec 16, 1999 this sequence version replaced gi:5763807.
            For more information about this sequence or the Malaria Project,
            see http://www.sanger.ac.uk/Projects/P_falciparum. IMPORTANT: This
            sequence is unfinished and does not necessarily represent the
            correct sequence. Work on the sequence is in progress and the
            release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc.

FEATURES
    source
        location/Qualifiers
            1..67970
            /organism="Plasmodium falciparum"
            /strain="3D7"
            /db_xref="taxon:5833"
            /chromosome="1"
            /gene="MALIP3.01"
            /complement(join(1748..2598,2748..2848,2990..3276))
            /gene="MALIP3.01"
            /complement(join(1748..2598,2748..2848,2990..3276))
            /gene="MALIP3.01"
            /note="MALIP3.01, conserved hypothetical protein, len: 412
            aa, similarity: UF0006 family eg to
            YBL055C/YBL0512/YBL0511, YB5_YEAST (418 aa), fast
            scores: opt: 316, E(): 1.1e-12, (33.2% identity in 271 aa
            overlap)"
            /codon_start=1
            /product="conserved hypothetical protein, UPP0006 family"
            /protein_id="CAB63556.1"
            /db_xref="GI:6594244"
            /translation="MKLVFHYIKYINLVFYISIFLKSNSLKIYNDLRYSTVNVKKY
            LQIKRSNLKKNHNIKMEVNESSFIDIGSLTLTKMPGYNVSKKHENDLQNVLRK
            NNNDKILITCTCLAEIDSKICETYPDEKPTPLYSAGVPTNCYETPIDKKEHEK
            ILAKREIEFTIKRNEOVENSKEKNGKKICDCEKDMNNNEILLENLDTIFGFKY
            NEKDEYLENFKNKIKIPNRIIVGIGELGIDRLYFCSKYIQIKYFLOLVQMFN
            LPMLHMRNCSEFEKIVDIYKFLPEKNGGVYHSFTQEDYIVHIVQYKNLYGVNG
            CSLKSLKLEINAVKIKIPNLILLETDAPCGVKTHASVEYIKDYEKRAYTNLKKIKN
            IIKDDNTIFKERNRPYNA"
            /complement(2599..2610)
            /note="potential splice acceptor sequence"
            /complement(2742..2747)
            /note="potential splice donor sequence, atg/gttaa"
            /complement(2849..2861)
            /note="potential splice acceptor sequence"
            /complement(2984..2989)
            /note="potential splice donor sequence, aaa/gtaaaa"
            /gene="MALIP3.02"
            /gene="MALIP3.02"
            /note="MALIP3.02, hypothetical protein, len: 163 aa,
            contains possible signal sequence"
            /product="hypothetical protein, MALIP3.02"
            /codon_start=1
            /protein_id="CAB63557.1"
            /db_xref="GI:6594245"
            /translation="MKLLNRFVVLCPITIIFFLNSVGLGNRRNNINHEHETNAK
            AMRLLGGEINSLKLDNGDEIKTLNDEKHKDYSFKWKSYSFISNLEEKQSQIDLR
            KQDINENATKIIDROEFYILNDEIENIATREVLNNDELYIOSFKSLDIIOS

```





	repeat_unit	/organism="Drosophila melanogaster" /organism="nullochondria" /strain="Oregon R" /db_xref="taxon:7227" /dev_stage="embryo" 650..1022 /note="repeat I-A"
	repeat_unit	/rpt_type=tandem 1023..1360 /note="repeat I-B1" /rpt_type=tandem 1361..1705 /note="repeat I-C/A"
	repeat_unit	/rpt_type=tandem 1706..2043 /note="repeat I-B2" /rpt_type=tandem 2044..2388 /note="repeat I-C"
	repeat_unit	/rpt_type=tandem 2491..2511 /note="deoxythymidylate stretch" 2512..2648 /partial
	repeat_unit	/rpt_type=tandem 2649..3112 /note="repeat II-A" /rpt_type=tandem 3113..3576 /note="repeat II-B1"
	repeat_unit	/rpt_type=tandem 3577..4040 /note="repeat II-B2" /rpt_type=tandem 4041..4504 /note="repeat II-C"
	misc_feature	/rpt_type=tandem complement(4565)..4585) /note="deoxythymidylate stretch"
BASE COUNT	2271 a	131 c 74 g 2125 t
ORIGIN		

	Matches	1045;	Conservative	0;	Mismatches	1284;	Indels	42;	Gaps	11;	
OY	83	tttggttc	caattggatccgaagaa	ttaata	ltaaata	lcatltaa	aatttacc	142			
Dd	3827	TTAAATTTTAAAAA	TTTTTTTATTATAATTAATATGATATTAATATTTAT					3768			
OY	143	aagcaaat	aatccttaccatg	tatctgtat	gtttaacaaa	atctcatc	cttggatalga	202			
Dd	3767	ATATATATAATACAT	TCTAT-----TATTTATATATAGTATGTATGTTTTTTTTTAAAAA					3714			
OY	203	gaacaat	aggagtctggaat	ltaataa	laagaga	aacacga	lctcatctggttga	262			
Dd	3713	AAAATTTATTTTTTTT	TTAAAAAATTTTTTAAAAAAATTTGAAGAAATTAATATATATTTCA					3654			
OY	263	tttaacag	gttaagtttlttgtgtctcttgc	tatalgbtal	lgag	ltaacaaa	agagt	322			
Dd	3653	TTATATAATTTATTTT	TTAAAAATTTTTGGT-----TATTTTTTAAAAAACATGATTT					3601			
OY	323	atgatctga	agtglaaaca	latctglta	tgccccca	aaaaa	aaaaaacaacaa	382			
Dd	3600	TTATATATATAATTTT	TATAAAAATATATCATTTAGAATTTTAAAAAATTTATA					3541			
OY	383	acaaacccc	cccccgatatg	tttggctc	ggatta	aggtlta	atttgacataat	442			
Dd	3540	TTAAATTTATTTAAAT	AATTTAAATTTTTTCATATATATATATATATATATATATATCA					3481			
OY	443	atgcata	cttcttgat	tactaaga	agattctt	caccata	aaatttcgaatcata	502			
Dd	3480	TAAATATATAATTTAT	TAATAATATATATATATATATATATATATATATATATATAT					3421			
OY	503	tccctg	atattaa	atlaa	cagagtg	gaata	lccgttat	cgalccatccatcat	562		
Dd	3420	TAAATTAATATAATAA	TAAATAATGAGAAATATAAT-----TTTATAAATTTATCTGCATAT					3365			
OY	563	gata	gatctctg	tgcbtac	ccagca	aatatt	taaca	aggtltagagaaaaaccga	622		
Dd	3364	TTTAAATTTTAAAA	TTTTTTTATTATTTATAGTATATATATATATATATATATATATAT					3305			
OY	623	aataa	gaaagggaag	tagtga	cccatbg	aglatg	atataat	ltaacagaaga	682		
Dd	3304	TATATATATAAATATC	TATTAATTAATAGTATATATATATATATATATATATATAT					3245			
OY	683	agsga	gacaccca	aaagggtg	tgga	aaatag	gtccctgc	agcttctctcaacalca	742		
Dd	3244	TATTTTTTTTAAAA	AAATTTTTTAAAAAATTTGAAAAAATAATATATATATTTCCATTATA					3185			
OY	743	falcga	cccatl	tgatctt	ctcgatal	btcgtt	aaattcg	caataagat	792		
Dd	3184	AAATTTATTTTAAAA	ATTTTTTGTTTATTTTTTAAAAAACATGATTTTATATATATATA					3125			
OY	803	atactt	ltaattg	tgcgc	gatc	caata	atlatg	tlcca	aggtatctg	862	
Dd	3124	TATTTTTTTTAAAA	AAATTAACATTTTAGAAATTTTTTAAAAAATTTTATTAATTAATTTTA					3065			
OY	863	taaga	catatt	ctgtctt	ctctt	gttctgt	cttcc	taacata	latalat	922	
Dd	3064	AATTAATTTAATTTTT	CTAT					3005			
OY	923	fattga	acaatgat	ata	ltaacaa	aatlgt	ctgg	accattl	gaa	982	
Dd	3004	TTATATAATATATATA	TAATTAATTTTAAATTTTATATAT-TATTTATATAATTTAAATTAATA					2946			
OY	983	tcaac	acatl	agggac	ctgac	cgccc--	ctta	aatcga	ttlta	1040	
Dd	2945	TAAATAATATGAGAA	NATATAATTTTATATAATTTATATCTACTTTTTTAAATTTTAAAT					2886			
OY	1041	ttgag	att	actag	aca	ataa	gacac	cgctt	caagct	1100	
Dd	2885	TTTTTATTTAATTTAT	TAGATATATATAATAATATATATAATTTTATATATATATATATAAT					2826			
OY	1101	ctga	at	atag	ctt	gaaat	ctctt	gaa	attacc	1160	
Dd	2825	CTATATATTTATATA	TAT					2766			

Oy	1161	gatttgatggtgaaacaaatctgttaactcaatctgtgctgaacaaatctcaactcttacc	1220
Db	2765	ATTTTTTTTTTAAAAATGAAAAATAAATTAATTAATTTTATTATTAATAATTTATTATTTAA	2706
Oy	1221	agaaaaaacaacttaagagaagttlaacatlatccatactggtgatacgtctaccccttacc	1280
Db	2705	AAATTTTTTTTGTTTTTTTTTAAAAAACATGATTTTTTATATATAAATATTTTTTATAAAAA	2646
Oy	1281	gtatgctactactgagagcttaagaatcagttatgtgtgtgcgtatgaatgaatccacgc	1340
Db	2645	TAAATACATTTAAGAAATTTTTTAAAAAATTTATATTAATAATTTATTAATTTAAATTTTTT	2586
Oy	1341	gtggtaactaaactatgaggaaccgtiatgtctagatcaactcgcaatcatccttctgttgca	1400
Db	2585	CTATATATATATATATATTTATTAATATATTCATATATATTAATTTATTAATAATATATAA	2526
Oy	1401	acaataaacaacaaacacagaacaaagaacagatcttcttgatctccatccaagat	1460
Db	2525	TTAATTTAAATTTATTTAAAAAATAAAAAAATAAATTTTATTTATTTATTAATAATAT	2466
Oy	1461	ctaaatgcatagactcttcttggtgtacagtttcaggaagtcctctacaaagcgigtaccc	1520
Db	2465	TAGTATATTAATAATTTTTATTATTAATTAATTAATTAATTTATTTATTTATTAATTAAT	2406
Oy	1521	tgcacattaaactgtcctcttcttaactgcaacttcaacatatactgtctgtagat	1580
Db	2405	ATTATATATGAAATGATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAAAT	2346
Oy	1581	taataa-gagsgaactgttaacttccaatatttatac-tagatactgtagtggatcat	1638
Db	2345	TAAAAATTTATTTTCATTTTATATATATATATATATATATATATATATATTTATTTAAAT	2286
Oy	1639	lccaacata-----catccttggagctttaaactaactctgtctctccocacgta	1690
Db	2285	TTTTAAATATTTTATTTATTAATAATTAATTTATTAATAAATAGTTATTAAGATATAATTA	2226
Oy	1691	taaatatatactacgagaggttaaaaaaagtttgcctatcttcgcgagatgaagaata	1750
Db	2225	ATTAATTAATTTTTTTTTTAAAAAATAATTTTTTTTAAAGTTTTATTAATTAATTAATAATTA	2166
Oy	1751	aacctaaactttaaactcttcttgaanaatgtaacccttaccacagataactaacgta	1810
Db	2165	TGATATAGGGGGAATAATTTATTTTCATTTTTCATATATATATATATATATATATATATCA	2106
Oy	1811	tgcttctgtgcctaaatgcagagccctcaacactgtagatagtaactcttctcgcacata	1870
Db	2105	ATTATATTAATTTACATTTTACGTATTAATAATTAATAATTTATTAATTAATCAATATATTTAA	2046
Oy	1871	ctaaacttagaatcaactgtctacttccaatagaagaacagctgtgatacttcaattat	1930
Db	2045	TTGAAATTTAATATATATATGATATATATATTAATTAATAATATATCAATTTGA--ATTTTTTAAAAA	1968
Oy	1931	ctaaagacaanaatttggaaaatgltataaattctbaacatatataatnaaatalgagc	1990
Db	1987	ATCATTTTAAATTTTTTTTATTTATTTAAAAAATTTTTTATTAATTTAAAAATATTTTAT	1928
Oy	1991	ctctaaatgcttccactgtctctctaaatattttttatacttagtataaataact	2030
Db	1927	TTTATTAATAATTTATTTATTTATTAATAAATAGTTTATTAAGTATAATTTATTAATAATCATTTTTT	1868
Oy	2051	atgaaaccaataatagctgtgtagaatccaacatcccatataattcttgaactctaca	2110
Db	1867	TTTAAAAAATAATTTTAAAAAGTTTATTAATTTATTAATAATTTTAAAAAGTTTAA--TTTATGAATGGGGGA	1811
Oy	2111	attatataatttagctcaataaacaatgcatagaaagttcccaaaaaaaattctgttaacag	2170
Db	1810	ATTAATTTATTTTTCATTTTATTTTATTAATATATATATATATATATTAATTTATTTTCAGATT	1751
Oy	2171	aaacttccaactttcttcttcttctgaagcaagaataacagatagaagaacatcttctt	2230
Db	1750	TAGCATATTAATAATTAATTTATTTATTTATTTATTTATTAATTTATTTATTTATTTAAATTTAAATTT--1692	

Oy	2231	gtggagatggaagtagtaatacatcaagaacaaattttaaataatatagccctacgc	2290
Dd	1693	-----TATATATATATATATATATATATAAAATATTAATGCAATTTTTAAAAAATTAATTTTTAA	1638
Oy	2291	cgcctcaagatcgttatcctagtagaggltgatcaataatgcatggtcgatccaagtgcg	2350
Dd	1637	TTTTTATATATATAATAAAAATATTTCTTTATTAATTAATTTTAAAAATTTTATTAATAAATA	1578
Oy	2351	gacaacatggaacacgggaattcaaatatctaactttaaaaataaataaatttgtagaat	2410
Dd	1577	ATTATTTATATAAAATAAGTGTATTATTAAGTATATTTAATTAATAATTAATTTTAAAAAAA	1518
Oy	2411	gtgtttctgcacctatgagggcgcaaaaaa	2441
Dd	1517	ATATTTTAAAGTTTAAAGTTTAAATATATATAATRAA	1487
RESULT	14		
DMU37541/c			
LOCUS	DMU37541	19517 bp	DNA circular INV 02-MAR-2001
DEFINITION	Drosophila melanogaster complete mitochondrial genome.		
ACCESSION	U37541		
VERSION	U37541.1	GI:1166529	
KEYWORDS			
SOURCE	fruit fly.		
ORGANISM	Mitochondrion Drosophila melanogaster		
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
AUTHORS	Clarry,D.O., Goddard,J.M., Martin,S.C., Fauron,C.M. and Wolstenholme,D.R.		
TITLE	Drosophila mitochondrial DNA: a novel gene order		
JOURNAL MEDLINE	Nucleic Acids Res. 10 (21), 6619-6637 (1982)		
REFERENCE	83090428		
AUTHORS	2 (bases 5269 to 5695) Clarry,D.O., Wahleithner,J.A. and Wolstenholme,D.R.		
TITLE	Transfer RNA genes in Drosophila mitochondrial DNA: related 5' flanking sequences and comparisons to mammalian mitochondrial tRNA genes		
JOURNAL MEDLINE	Nucleic Acids Res. 11 (8), 2411-2425 (1983)		
REFERENCE	83220794		
AUTHORS	3 (bases 404 to 5272) de Bruijn,M.H.		
TITLE	Drosophila melanogaster mitochondrial DNA, a novel organization and genetic code		
JOURNAL MEDLINE	Nature 304 (5923), 234-241 (1983)		
REFERENCE	83245048		
AUTHORS	4 (bases 804 to 1778) Satta,Y., Ishiwa,H. and Chigusa,S.I.		
TITLE	Analysis of nucleotide substitutions of mitochondrial DNAs in Drosophila melanogaster and its sibling species		
JOURNAL MEDLINE	Mol. Biol. Evol. 4 (6), 638-650 (1987)		
REFERENCE	88174373		
AUTHORS	5 (bases 5268 to 13619) Garesse,R.		
TITLE	Drosophila melanogaster mitochondrial DNA: gene organization and evolutionary considerations		
JOURNAL MEDLINE	Genetics 118 (4), 649-663 (1988)		
REFERENCE	88212147		
AUTHORS	6 (bases 441 to 2967) Satta,Y. and Takahata,N.		
TITLE	Evolution of Drosophila mitochondrial DNA and the history of the melanogaster subgroup		
JOURNAL MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 87 (24), 9558-9562 (1990)		
REFERENCE	91088557		
AUTHORS	7 (bases 14215 to 14512) Ballard,J.W., Olsen,G.J., Faith,D.P., Odgers,W.A., Rowell,D.M. and Atkinson,P.W.		
TITLE	Evidence from 12S ribosomal RNA sequences that onychophorans are modified arthropods		
JOURNAL MEDLINE	Science 258 (5086), 1345-1348 (1992)		
REFERENCE	93088057		



```

REFERENCE      8 (bases 14917 to 19517)
AUTHORS        Lewis,D.L., Farr,C.L., Farquhar,A.L. and Kaguni,L.S.
TITLE          Sequence, organization, and evolution of the A+T region of
               Drosophila melanogaster mitochondrial DNA
JOURNAL        Mol. Biol. Evol. 11 (3), 523-538 (1994)
MEDLINE        94285822
REFERENCE      9 (bases 1 to 408; 13319 to 19517)
AUTHORS        Lewis,D.L., Farr,C.L. and Kaguni,L.S.
TITLE          Drosophila melanogaster mitochondrial DNA: completion of the
               nucleotide sequence and evolutionary comparisons
JOURNAL        Insect Mol. Biol. 4 (4), 263-278 (1995)
MEDLINE        96423163
REFERENCE      10 (bases 1 to 19517)
AUTHORS        Lewis,D.L., Farr,C.L. and Kaguni,L.S.
TITLE          Direct Submission
JOURNAL        Submitted (03-OCT-1995) Laurie S. Kaguni, Biochemistry Department,
               Michigan State University, East Lansing, MI 48824-1319, USA
FEATURES       source
               1. 19517
               /organism="Drosophila melanogaster"
               /organelle="mitochondrion"
               /db_xref="taxon:7227"
               /note="derived from new and previously submitted
               sequences; sequence is a composite containing sequences
               obtained from different Drosophila melanogaster strains"
               1. 65
               /product="tRNA-Ile"
               complement(97..165)
               /product="tRNA-Gln"
               171..239
               /product="tRNA-Phe"
               240..1265
               /codon_start=1
               /transl_table=5
               /product="NADH dehydrogenase subunit 2"
               /protein_id="AAC47811.1"
               /db_xref="GI:1166530"
               /translation="MFNNSSKLFITIMIGLITVTSNMGAMGELINILSPIL
               LSDNNNMSTESAKFQVOYASTVYLESSILMLKNNMNEINSPISMITMSLL
               LKSGAPFHPFNPMEGLITMNAALMTWKIAPMLISYNIKILILSVLSTI
               GAIIGLNOTSLRKLKLAFFSSINHLGWLSSIMISESILILFEFFSFLSVLFMEFIF
               KLFHLNOLFESVNSKILKFTLFMNFSLISGGLPPFLGFLPKMLVIOQLTLCQYFWLT
               IMMSTLITLFFYLRCYSAPFMNVEFNMMIKMNNNSINYMMYIMTFEFSIFGLFLI
               SLFVFMF"
               1264..1329
               /product="tRNA-Trp"
               complement(1322..1383)
               /product="tRNA-Cys"
               complement(1403..1468)
               /product="tRNA-Tyr"
               join(1470..1472,1474..3009)
               /note="mechanism underlying reading frame shift after
               first codon uncertain"
               /codon_start=1
               /transl_table=5
               /product="cytochrome c oxidase subunit I"
               /protein_id="AAC47812.2"
               /db_xref="GI:7412849"
               /translation="MSRQWLFSFNHNDIGLYTIFGAMGMYTSLILRAELGHPG
               ALIGDQIYNYIVTAHAFIMIFPMWPIGIGFAMLVPLMGAPMAPPRMNMSEFM
               LIPALSLILSVNENGAGTGMWYVPPISAGIAGAGAVDLAIFSLHAGISLIGA
               VNFITVIMRSTGTSIDRMLPVMGVSVTALLLLSLPLAGATIMLTDRLNISF
               FPPAGGGPILIOHLFHPFGHEVYILILPGFMISHIISQESGKETGSGIMTYAM
               LAIIGLGTVMAHMFTVGMVDYTRAFYTSATMIJAVPGIKIFSWLATLHGTOISY
               PAILMAIGFVLTGVLGVLANSSVDIILHDYVYVAHRYVLSMGAVPAIMAGF
               IMHYPLFTGLTNKMLKSHFTIMFGVNLITFEPQHFLLAGMPRYSYDPDAYTTW
               IYSTGTSISLIGILFEFFIIMESLVSQROYVPIQLNSIEMYQMPAHSYSFLP
               ILTN"
               3012..3077
               /product="tRNA-Leu"
               3083..3767
               /note="TAA stop codon is completed by the addition of 3' A
               residues to the mRNA"

```

```

               /codon_start=1
               /transl_except=(pos:3767,aa:TERM)
               /transl_table=5
               /product="cytochrome c oxidase subunit II"
               /protein_id="AAC47813.1"
               /db_xref="GI:1166532"
               /translation="MSTWANTLGLDASPLAMEQLIFPHDALLITVITVLGYLWM
               LFENNVNRFILGOLIEIMITLIPAILFTALPLSLRLYLDEINSPVTLKSI
               GHSYKSEYSDENNIEFDSYMIPTNLMDDGPELLDVDRNYVPMNSQIRIITVADYI
               HSWTPALGVKVDGTPGRUNOTNFTNREGLFYGGCSEICGANHSMPVIESVYNT
               FIKWISSNNS"
               3768..3838
               /product="tRNA-Lys"
               3840..3906
               /product="tRNA-Asp"
               3907..4068
               /codon_start=1
               /transl_table=5
               /product="Argase 8"
               /protein_id="AAC47814.1"
               /db_xref="GI:1166533"
               /translation="MPQAPISMLLFIITFIFLFCISINYSTMPSKSNELKI
               NLSMNWKW"
               4062..4736
               /codon_start=1
               /transl_table=5
               /product="Argase 6"
               /protein_id="AAC47815.1"
               /db_xref="GI:1166534"
               /translation="MWTNLFVYFDPLAIFNFSLNLSTFLGLMTPSIYWLMPRYNI
               MNNSILTLTKKEKFTLGPSSNGNSSTFIFISLSLILFNNMGLPEYIFTSHTLT
               LSLALPILCMPLMGWINHTOHMFAHLVPOGPAILMPMVCETISNIIIRGTLAVR
               LTNMAGHLITFLIGNTSSSYMLMTLMAQIALVLVESVAMIGSYVAVASTL
               YSEVN"
               4736..5524
               /codon_start=1
               /transl_table=5
               /product="cytochrome c oxidase subunit III"
               /protein_id="AAC47816.1"
               /db_xref="GI:1166535"
               /translation="MSTHSNHPFHLVDYSPWPLTGAIGAMTYVSGWVKWPHOYDISL
               FVGNITITIVYQWRDVSREGTYOGLHYAVTIGLRGMILFIISEVLFVFSFWAF
               FHSLSPLALEIGSPMPMGITSPNPQIPLMTALILASGVYTAHHSILMENNHSOT
               TQGLFPTVLTGYFTTIOAYEYIEAPETIADSIYSGTFPMAGFGHGHVILGTFILY
               CLRLHNHFSKNHFGFEDAAAMYHFDVYVWVLFYITTYWNG"
               5543..5607
               /product="tRNA-Gly"
               5608..5961
               /codon_start=1
               /transl_table=5
               /product="NADH dehydrogenase subunit 3"
               /protein_id="AAC47817.1"
               /db_xref="GI:1166536"
               /translation="MFSIIFIALILLITTYIVFLASILSKKALIDREKSSPECGFD
               PKSSRLPSLRPLFTITLIFDFVEIALILPMLIIMKYSNIMIMWTISIFILLILI
               GLYHENGCMKMSN"
               5961..6045
               /product="tRNA-Ala"
               6055..6118
               /product="tRNA-Arg"
               6119..6183
               /product="tRNA-Asn"
               6184..6251
               /product="tRNA-Ser"
               6252..6318
               /product="tRNA-Glu"
               complement(6337..6401)
               /product="tRNA-Phe"
               complement(6401..8124)
               /note="TAA stop codon is completed by the addition of 3' A
               residue to the mRNA"
               /codon_start=1
               /transl_except=(pos:complement(6401..6402),aa:TERM)

```







[illegible]

Db	10000	AATATTATATATTTTAAATATAAATTAATATATTTTACTATTTAAATATAATTAATTTTA	10059
Qy	2189	ttttatcggaacagaataaacacagatagaanaacttttgctgcyggaatagggaagcagtaa	2248
Db	10060	TATTTTAAATTAATAATATATTTTAAATATAAATTAATATTTTTCATATATTTAAATTT	10119
Qy	2249	tatcaatgaagcaaattttaaaaatatataagccatacagcgctcaagaatgtatc	2308
Db	10120	AATTAATTAATAATATTTTATATATGCAATATATTTTATATAATTAATTAATTTATTTATG	10179
Qy	2309	tagtaggtgtaattcataatgcacgtgctgcgaattcggaattgggacaacaatgaaacgga	2368
Db	10180	TATTTATATTTTATTAATAATTTATTTTGTGTGTATAATTTAATTAAGATTTTATATA	10239
Qy	2369	atcaaatatctactttaataaataaataa 2398	
Db	10240	TTGATGTTTATTTTATATAATTAATAATTAATA 10269	

Search completed: October 17, 2001, 19:26:38  
Job time: 16993 sec

Oy 1781 aaccttctactcaaatatcacgaatgcttcttgcccaatgaagccctcac 1840  
 Db 9640 TTAATATGTAATTAATTAATAATTAATTTTATTTATATATTAATATATTTTAT 9699  
 Oy 1841 aactgtagtcaattttctgcacaatataatlaagaaatcaagctactaca-a 1899  
 Db 9700 ATAAATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9759  
 Oy 1900 tagaagaacagcgagatcaacttaatttaaaagcaaaatlttggaaaagtlat 1959  
 Db 9760 TTTATTTATTAATCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 9819  
 Oy 1960 aattt---ctaacacatctatlaaatalagatgccataatglatccat----- 2008  
 Db 9820 ATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9879  
 Oy 2009 gtctctaaaatlttttctatctagtgtlaaataacatcagaaaccaataagctg 2068  
 Db 9880 ATATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9939  
 Oy 2069 gtgaatcacaatctccacttaatttctgaaactcaacaattataatctagctca 2138  
 Db 9940 ATATTAATTAATTAATTAATTAATTAATTAATTTCTTTTAAATTAATTAATTAATTA 9999  
 Oy 2129 atacaatgcatagaaagctccaaaaaaatttgttcaagaaacttccaaatttttt 2188



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: October 17, 2001, 14:53:20 ; Search time 431.79 Seconds  
(without alignments) 4656.294 Million cell updates/sec

```

Title:      US-09-502-426-1_COPY_1_3202
Perfect score: 3202
Sequence:   1 atgtggtattatatgttg.....gagagagaaactagctcc 3202

```

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

```
Searched:      730101 seqs, 313950809 residues
Total number of hits satisfying chosen parameters: 14602022
```

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

Listing first 45 summaries

1:	/SIDSL/gcgdata/genseq/genseqn/NA1960.DAT *
2:	/SIDSL/gcgdata/genseq/genseqn/NA1961.DAT *
3:	/SIDSL/gcgdata/genseq/genseqn/NA1962.DAT *
4:	/SIDSL/gcgdata/genseq/genseqn/NA1963.DAT *
5:	/SIDSL/gcgdata/genseq/genseqn/NA1964.DAT *
6:	/SIDSL/gcgdata/genseq/genseqn/NA1965.DAT *
7:	/SIDSL/gcgdata/genseq/genseqn/NA1966.DAT *
8:	/SIDSL/gcgdata/genseq/genseqn/NA1967.DAT *
9:	/SIDSL/gcgdata/genseq/genseqn/NA1968.DAT *
10:	/SIDSL/gcgdata/genseq/genseqn/NA1969.DAT *
11:	/SIDSL/gcgdata/genseq/genseqn/NA1990.DAT *
12:	/SIDSL/gcgdata/genseq/genseqn/NA1991.DAT *
13:	/SIDSL/gcgdata/genseq/genseqn/NA1992.DAT *
14:	/SIDSL/gcgdata/genseq/genseqn/NA1993.DAT *
15:	/SIDSL/gcgdata/genseq/genseqn/NA1994.DAT *
16:	/SIDSL/gcgdata/genseq/genseqn/NA1995.DAT *
17:	/SIDSL/gcgdata/genseq/genseqn/NA1996.DAT *
18:	/SIDSL/gcgdata/genseq/genseqn/NA1997.DAT *
19:	/SIDSL/gcgdata/genseq/genseqn/NA1998.DAT *
20:	/SIDSL/gcgdata/genseq/genseqn/NA1999.DAT *
21:	/SIDSL/gcgdata/genseq/genseqn/NA2000.DAT *
22:	/SIDSL/gcgdata/genseq/genseqn/NA2001.DAT *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3202	100.0	6888	21	AAF59599	DNA encoding a cyto
c	251	7.8	936	22	AAAF58252	Oligonucleotide D1
c	251	7.8	936	22	AAAF58254	Oligonucleotide D1
c	251	7.8	936	22	AAAF58257	Oligonucleotide D1
c	251	7.8	936	22	AAAF58259	Oligonucleotide D1
c	251	7.8	936	22	AAAF58262	Oligonucleotide D2
c	251	7.8	938	22	AAAF58255	Oligonucleotide D1
8	250.6	7.8	936	22	AAAF58252	Oligonucleotide D1
9	250.6	7.8	936	22	AAAF58254	Oligonucleotide D1
10	250.6	7.8	936	22	AAAF58257	Oligonucleotide D1
11	250.6	7.8	936	22	AAAF58259	Oligonucleotide D2

C	19	74	7.4	2.3	2104	13	AAO25273	Sequence encoding
C	20	73	7.3	2.3	1864	8	AAW1405	Sequence encoding
C	21	70.8	70.8	2.2	8310	20	AAZ29911	cDNA encoding a SC
C	22	69.6	69.6	2.2	4590	7	AAW0472	Sequence encoding
C	23	68.8	68.8	2.1	875	21	AAAO1920	Human colon cancer
C	24	68.6	68.6	2.1	1864	8	AAW1405	Sequence of ANS-1
C	25	68.2	68.2	2.1	2104	13	AAO25273	Sequence encoding
C	26	67.4	67.4	2.1	5760	6	AAW50530	Sequence encoding
C	27	67.4	67.4	2.1	19124	18	AAW27882	Plasmodium var-7 g
C	28	67.4	67.4	2.1	19124	21	AAZ96287	Plasmodium var-7 p
C	29	67.2	67.2	2.1	875	21	AAAO1920	Human colon cancer
C	30	66	66	2.1	2486	21	AAW97037	Nucleotide sequence
C	31	65.8	65.8	2.1	2486	21	AAW97037	Nucleotide sequence
C	32	65.4	65.4	2.0	2593	15	AAO53480	pnpx30 xylanase cd
C	33	65.4	65.4	2.0	20674	21	AAO58017	Arachnid acid m
C	34	65.2	65.2	2.0	6033	21	AAW70152	Plasmodium falcipea
C	35	65	65	2.0	2418	13	AAO27886	p.falciparum GBPI3A
C	36	64.4	64.4	2.0	5409	21	AAW70151	Plasmodium falcipea
C	37	64.2	64.2	2.0	1431	21	AAZ37082	DNA sequence encod
C	38	64.2	64.2	2.0	1671	13	AAO24134	50 kb subunit of S
C	39	63.6	63.6	2.0	8920	15	AAO62824	Carbamoyl-phosphat
C	40	63	63	2.0	20674	21	AAO58017	Arachnid acid m
C	41	62.6	62.6	2.0	4818	18	AAW62177	Candida boidinii A1
C	42	62.6	62.6	2.0	4818	18	AAW79909	Actin gene. Candid
C	43	62.6	62.6	2.0	9789	17	AAW41852	cDNA encoding Plas
C	44	61.8	61.8	1.9	1864	15	AAW78892	Aspergillus nidulans
C	45	61.8	61.8	1.9	5852	12	AAQ11710	Dicoryostellum plas

## ALIGNMENTS

## RESULT 1

ID AAA59599 standard; DNA; 6888 BP.

AC AAA59599

DT 14-NOV-2000 (first entry)

DE DNA encoding a cytochrome P450 enzyme designated DMF4

KM DWPF4; cytochrome P450 enzyme; brassinosteroid; 22alpha-hydroxylation  
KM plant phenotype; cell elongation; ss.

OS *Arabidopsis* sp.

Key	Location/qualifiers
FH	1 3303

FBI  
FBI  
FBI

ET  
ET  
CNC

32

FT  
exon

FT intron

FT exon

FT intron

FT exon

```
FT      Intron      4067..4164      /*tag= 1
FT      Exon        4165..4479      /*tag= j
FT      Intron      4480..4631      /*tag= k
FT      Exon        4632..4724      /*tag= l
FT      Intron      4725..4815      /*tag= m
FT      Exon        4816..4894      /*tag= n
FT      Intron      4895..5000      /*tag= o
FT      Exon        5001..5110      /*tag= p
FT      Intron      5111..5864      /*tag= q
FT      Exon        5865..6110      /*tag= r
FT      3'UTR       6011..6468      /*tag= s
XX      WO200047715-A2.
XX      17-AUG-2000.
XX      11-FEB-2000; 2000WO-US03820.
XX      11-FEB-1999; 99US-0119657.
XX      11-FEB-1999; 99US-0119658.
XX      (ARIZ-) ARIZONA BOARD OF REGENTS.
XX      Azpiroz R, Choe S, Feldmann KA;
XX      WPI; 2000-549142/50.
XX      P-PSDB; AAB07921.
XX      New isolated dwf4 polynucleotide useful for altering the phenotype of
XX      plants, for diagnostic assays and in the production of antibodies -
XX      Claim 3; Fig 10A-G; 113pp; English.
XX      The present sequence encodes a DWF4 polypeptide. The polypeptide is a
XX      cytochrome P450 enzyme that mediates multiple steps in synthesis of
XX      brassinosteroids. Specifically, it mediates multiple
XX      22alpha-hydroxylation steps in brassinosteroid biosynthesis. The DWF4
XX      polynucleotide is used for altering the phenotype of a plant. DWF4
XX      plants display a dramatic reduction in the length of different organs,
XX      and this size reduction is attributable to a defect in cell elongation.
XX      The DWF4 polynucleotides and polypeptides can be used in diagnostic
XX      assays and to generate antibodies, which can be used to produce
XX      immunogenic compositions.
XX      Sequence 6888 BP; 2294 A; 1010 C; 1193 G; 2391 T; 0 other;
SQ
Query Match      100.0%; Score 3202; DB 21; Length 6888;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3202; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      181 aatattcatttggatataagaataataggatttgaatttaataataaaggaa 240
DB      181 aatatcattcttggatataagaaaaataggatttgaatttaataataaaggaa 240
QY      241 taatcattccatttggattgattacacagttlaagtttggttcttcttggatattg 300
DB      241 taatcattccatttggattgattacacagttlaagtttggttcttcttggatattg 300
QY      301 atattgtaaatcaaaaagagattgattgaagtgtaaacatatcttcgttatgacccca 360
DB      301 atattgtaaatcaaaaagagattgattgaagtgtaaacatatcttcgttatgacccca 360
QY      361 aaaaaaaaaaaaaaacaacaaaccccccccccgatatagtttggttcgatt 420
DB      361 aaaaaaaaaaaaaaacaacaaaccccccccccgatatagtttggttcgatt 420
QY      421 aggtatttgatcacaatatacatgacatcttcttgattactatgaagatttcttac 480
DB      421 aggtatttgatcacaatatacatgacatcttcttgattactatgaagatttcttac 480
QY      481 caattaaatttcgaattcatalctcttgatttaattaaatacagatgtaatatcc 540
DB      481 caattaaatttcgaattcatalctcttgatttaattaaatacagatgtaatatcc 540
QY      541 gttatcgatccactccaatcatgatattatattcttgtaacacgaataattataca 600
DB      541 gttatcgatccactccaatcatgatattatattcttgtaacacgaataattataca 600
QY      601 agagatttgaagaaaaaacggaataataagaagaagagtagtaccatgagatag 660
DB      601 agagatttgaagaaaaaacggaataataagaagaagagtagtaccatgagatag 660
QY      661 tgaataattatcaaaagagataagagatgaacacaaaggttgtaataaggtccct 720
DB      661 tgaataattatcaaaagagataagagatgaacacaaaggttgtaataaggtccct 720
QY      721 gccagcttctctcaacaatcaatatacgaccctatttggatttcttgatatctgataaa 780
DB      721 gccagcttctctcaacaatcaatatacgaccctatttggatttcttgatatctgataaa 780
QY      781 ttgagataacgatttgtaaaaatatttatttgaagtcgatacatatattgtcca 840
DB      781 ttgagataacgatttgtaaaaatatttatttgaagtcgatacatatattgtcca 840
QY      841 ggaattgcaaatctctcgatttaagacatatttgcctcttcttggttctgctct 900
DB      841 ggaattgcaaatctctcgatttaagacatatttgcctcttcttggttctgctct 900
QY      901 taactatataatcgcgagatataatgaataacatgatatatacaaaacaattgtctgg 960
DB      901 taactatataatcgcgagatataatgaataacatgatatatacaaaacaattgtctgg 960
QY      961 accatttgaataaattttctcaacaatttcggaacactggaacgcgccttaataa 1020
DB      961 accatttgaataaattttctcaacaatttcggaacactggaacgcgccttaataa 1020
QY      1021 cgaatttacagctcactagttgagattactaacaataaagaccccgcttaacg 1080
DB      1021 cgaatttacagctcactagttgagattactaacaataaagaccccgcttaacg 1080
QY      1081 tattatacaaaagttaacaactgaatataagcttgaatacccttaagaanaatttgaatta 1140
DB      1081 tattatacaaaagttaacaactgaatataagcttgaatacccttaagaanaatttgaatta 1140
QY      1141 ccggtgttatgtgaaatataagatttagtggtaacaatatgttaataatgattgtgta 1200
DB      1141 ccggtgttatgtgaaatataagatttagtggtaacaatatgttaataatgattgtgta 1200
QY      1201 acatatacaatattccttcaagaaaaaacaacttaagaagttlaacatatccatata 1260
DB      1201 acatatacaatattccttcaagaaaaaacaacttaagaagttlaacatatccatata 1260
```



QY 1261 gggatgctataccttccacgtatgctactagagactaaagaatgtaatgtaatgctc 1320  
|||||  
Db 1261 gggatgctataccttccacgtatgctactagagactaaagaatgtaatgtaatgctc 1320  
QY 1321 gataaatgtaaatccacgcgtggtataataatgagaccgtatgtaacgatactgcaa 1380  
|||||  
Db 1321 gataaatgtaaatccacgcgtggtataataatgagaccgtatgtaacgatactgcaa 1380  
QY 1381 atatcatcttctggttgtaacaataaaaaaagaagaaaaaagaagatctttt 1440  
|||||  
Db 1381 atatcatcttctggttgtaacaataaaaaaagaagaaaaaagaagatctttt 1440  
QY 1441 ctgtgattccattcaatgtaactcaaaatgtaagatcttttgggttcaagtttcgaagtc 1500  
|||||  
Db 1441 ctgtgattccattcaatgtaactcaaaatgtaagatcttttgggttcaagtttcgaagtc 1500  
QY 1501 tctcaacgcgtgaacacatctgcacactataatgtcttcttcaatgcatcttcaacat 1560  
|||||  
Db 1501 tctcaacgcgtgaacacatctgcacactataatgtcttcttcaatgcatcttcaacat 1560  
QY 1561 atttatgttgaattgaatttaagaagcgaactgtgaacattacaataattatattag 1620  
|||||  
Db 1561 atttatgttgaattgaatttaagaagcgaactgtgaacattacaataattatattag 1620  
QY 1621 atactagtatgtatattccaaatacacacttggatgtttaaacttaactgtttct 1680  
|||||  
Db 1621 atactagtatgtatattccaaatacacacttggatgtttaaacttaactgtttct 1680  
QY 1681 tctctacggtataaatatcatcatcgaggttaaaaaagtttgccttatttcgcgagtc 1740  
|||||  
Db 1681 tctctacggtataaatatcatcatcgaggttaaaaaagtttgccttatttcgcgagtc 1740  
QY 1741 atgaagaataaacttaacttaacttaatttttgaanaatgtaacccttcaactatagatt 1800  
|||||  
Db 1741 atgaagaataaacttaacttaacttaatttttgaanaatgtaacccttcaactatagatt 1800  
QY 1801 aattacgcgtatgttttctgtgcataatgacagcctctacaactgtgatacttaatttt 1860  
|||||  
Db 1801 aattacgcgtatgttttctgtgcataatgacagcctctacaactgtgatacttaatttt 1860  
QY 1861 tctgcacaataatgaatgaatgcacgtctactatcaatagaagaacagcgtgattt 1920  
|||||  
Db 1861 tctgcacaataatgaatgaatgcacgtctactatcaatagaagaacagcgtgattt 1920  
QY 1921 accttttaatttaagaacaaatcttggaaaatgttatataattctcaacaattatttaa 1980  
|||||  
Db 1921 accttttaatttaagaacaaatcttggaaaatgttatataattctcaacaattatttaa 1980  
QY 1981 aatgatgacctataatgtaatttccatgtctcttaaaataatttttttatattagttta 2040  
|||||  
Db 1981 aatgatgacctataatgtaatttccatgtctcttaaaataatttttttatattagttta 2040  
QY 2041 taaatcatatgaacaaataatgattgtgtgaatcacaatctccatlaataattttttg 2100  
|||||  
Db 2041 taaatcatatgaacaaataatgattgtgtgaatcacaatctccatlaataattttttg 2100  
QY 2101 aaatctcaaaaattatattatattttagtcaataacaatgcatagaagaattccaaaaaaat 2160  
|||||  
Db 2101 aaatctcaaaaattatattatattttagtcaataacaatgcatagaagaattccaaaaaaat 2160  
QY 2161 ttgttaacagaactccaaattctttttttatgtgaacagaagaataacagatagaanaa 2220  
|||||  
Db 2161 ttgttaacagaactccaaattctttttttatgtgaacagaagaataacagatagaanaa 2220  
QY 2221 ctattttgtgtggaatggaatggaatgaataatataatagaacaaattttaaanaattatata 2280  
|||||  
Db 2221 ctattttgtgtggaatggaatggaatgaataatataatagaacaaattttaaanaattatata 2280  
QY 2281 agcctatacgcgtcacaagatgttatctcagtgatgtgtaataataatgatagtgtcgat 2340  
|||||  
Db 2281 agcctatacgcgtcacaagatgttatctcagtgatgtgtaataataatgatagtgtcgat 2340  
QY 2341 tcagaattggaacaacatgaaaaaggaatttaaatatttaacttaaatataataaat 2400  
|||||

Db 2341 tcagaattggaacaacatgaaaaaggaatttaaatatttaacttaaatataataaat 2400  
|||||  
QY 2401 ttgagttaaatgtgtttcttgcacatttgtaggggcaaaaaaagaacaaatgccaaagtctac 2460  
|||||  
Db 2401 ttgagttaaatgtgtttcttgcacatttgtaggggcaaaaaaagaacaaatgccaaagtctac 2460  
QY 2461 gggtttgactgcagttcggtaataatcctaactctgtcttgcacgcaagtcgtg 2520  
|||||  
Db 2461 gggtttgactgcagttcggtaataatcctaactctgtcttgcacgcaagtcgtg 2520  
QY 2521 taagggttccttgcacatttccactgttccacccactctgtgagccacccctttccca 2580  
|||||  
Db 2521 taagggttccttgcacatttccactgttccacccactctgtgagccacccctttccca 2580  
QY 2581 tatcctaagggttaatttttggaaaatcccaatttaacccgattgagccgtacccgtcc 2640  
|||||  
Db 2581 tatcctaagggttaatttttggaaaatcccaatttaacccgattgagccgtacccgtcc 2640  
QY 2641 tgggattctgcgtgagcatattatcaaaaatttatagacagaatgggtttatattatataa 2700  
|||||  
Db 2641 tgggattctgcgtgagcatattatcaaaaatttatagacagaatgggtttatattatataa 2700  
QY 2701 aactcaaacattgatcagataaatttcaataaacacttcttaacgattgattcgtacgactc 2760  
|||||  
Db 2701 aactcaaacattgatcagataaatttcaataaacacttcttaacgattgattcgtacgactc 2760  
QY 2761 atctaatgacttttttttctccacgcgtgtagtaagaattatagactattagccagag 2820  
|||||  
Db 2761 atctaatgacttttttttctccacgcgtgtagtaagaattatagactattagccagag 2820  
QY 2821 acaattgattatgataatataccaattcaatgataattatgataaataagctgttaa 2880  
|||||  
Db 2821 acaattgattatgataatataccaattcaatgataattatgataaataagctgttaa 2880  
QY 2881 ctatttcagcatcgacgcttctgcacacttcttcttaatttaagaagtttaataataa 2940  
|||||  
Db 2881 ctatttcagcatcgacgcttctgcacacttcttcttaatttaagaagtttaataataa 2940  
QY 2941 aagtattaaaaggacataaagcgaacaaagaatgaacagcgaagaaacaaagcca 3000  
|||||  
Db 2941 aagtattaaaaggacataaagcgaacaaagaatgaacagcgaagaaacaaagcca 3000  
QY 3001 tgaagctcatgtgttgaattgaagcttaataagaagatttatataattttaaagcagatg 3060  
|||||  
Db 3001 tgaagctcatgtgttgaattgaagcttaataagaagatttatataattttaaagcagatg 3060  
QY 3061 atacaattatatttctgtactctttaaaccctcttaacaacagagcctcccttt 3120  
|||||  
Db 3061 atacaattatatttctgtactctttaaaccctcttaacaacagagcctcccttt 3120  
QY 3121 tcagtagaagttccgattcccaactcttaaaagacaaagcattagaagaagaatgtgagta 3180  
|||||  
Db 3121 tcagtagaagttccgattcccaactcttaaaagacaaagcattagaagaagaatgtgagta 3180  
QY 3181 gagagagaggaactagctcc 3202  
|||||  
Db 3181 gagagagaggaactagctcc 3202

## RESULT 2

AAF58252/C

AAF58252 standard; DNA: 936 BP.

AAF58252;

24-APR-2001 (first entry)

Oligonucleotide D1835.

Electron-transfer group; ETM; mismatch; genotyping; gene expression; ss.

OS Synthetic.  
 XX WO200107665-A2.  
 PN  
 XX 01-FEB-2001.  
 PD  
 XX 26-JUL-2000; 2000WO-US20476.  
 XX PF  
 XX 26-JUL-1999; 99US-0145695.  
 PR 17-MAR-2000; 2000US-0190259.  
 XX  
 XX (CLIN-) CLINICAL MICRO SENSORS INC.  
 PA  
 XX  
 XX umek RM;  
 PI  
 XX  
 XX WPI: 2001-159728/16.  
 DR  
 XX  
 XX Nucleic acids containing electron-transfer group, useful as labels in  
 PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
 PT a single surface -  
 XX  
 XX  
 XX Example 6; Page 127; 159pp; English.  
 PS  
 XX The present invention relates to a composition comprising two nucleic  
 CC acids each containing an electron-transfer group (ETM) having  
 CC different redox potentials. The invention is used for electronic  
 CC detection of nucleic acids, especially of substitutions (mismatches)  
 CC and single-nucleotide polymorphisms, e.g. for genotyping,  
 CC monitoring gene expression.  
 CC  
 SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 7.8%; Score 251; DB 22; Length 936;  
 Best Local Similarity 1.0%; Pred. No. 7.7e-31;  
 Matches 8; Conservative 589; Mismatches 184; Indels 0; Gaps 0;

QY 1508 gcggtgaaccatcgcacatataatgcttctttaaagcatcttaacatattatt 1567  
 DB 793 GCGGCGCMMWW 734  
 QY 1568 gttagtgtgaatttaataagagcgaactgttaacatataattatttgatactag 1627  
 DB 733 WWW 674  
 QY 1628 tatgtattatccaatacatcacttgatgtttaaacttaacttcttctctaag 1687  
 DB 673 WWW 614  
 QY 1688 gataaataatcaatcagagtaaaaaagtttgcatttctcgagatgaag 1747  
 DB 613 WWW 554  
 QY 1748 ataaactaatgacttaatttttgaaaaatgtaacccttcaatcattatatacc 1807  
 DB 553 WWW 494  
 QY 1808 gtaatgttttgcataatgacagccctacaactgtgatgaatattttctgcaa 1867  
 DB 493 WWW 434  
 QY 1868 atataaataaggaatcgaatgtaatacaatagagaagaagctgagatatacttt 1927  
 DB 433 WWW 374  
 QY 1928 aatttaagaacaaattttgaaaaatgttaatttcaacataatttaaaataga 1987  
 DB 373 WWW 314  
 QY 1988 tgcctataatgataatcctcatgttcttaaatatttttataatttgatataaac 2047  
 DB 313 WWW 254

QY 2048 atatagaaccaataatagttgtaattcaaatatcattcaataattttgaaacta 2107  
 DB 253 WWW 194  
 QY 2108 caaatataataattagtcataacaaatgcatagaagttccaaaaaatttgtaa 2167  
 DB 193 WWW 134  
 QY 2168 cgaagaatccaaatttttttttaagacaagaataacagatagaataactatt 2227  
 DB 133 WWW 74  
 QY 2228 gtctgtgaatggaagtagtaataacataatgaagcaatttaaaaaatataagcctc 2287  
 DB 73 WWW 14  
 QY 2288 a 2288  
 DB 13 W 13

RESULT 3  
 ID AAF58254 standard; DNA; 936 BP.  
 XX AAF58254;  
 AC  
 XX  
 XX 24-APR-2001 (first entry)  
 DT  
 XX  
 XX Oligonucleotide D1875.  
 DE  
 XX  
 XX Electron-transfer group; ETM; mismatch; genotyping;  
 KW gene expression; ss.  
 XX  
 XX OS Synthetic.  
 XX WO200107665-A2.  
 PN  
 XX 01-FEB-2001.  
 PD  
 XX 26-JUL-2000; 2000WO-US20476.  
 PF  
 XX 26-JUL-1999; 99US-0145695.  
 PR 17-MAR-2000; 2000US-0190259.  
 XX  
 XX (CLIN-) CLINICAL MICRO SENSORS INC.  
 PA  
 XX  
 XX umek RM;  
 PI  
 XX  
 XX WPI: 2001-159728/16.  
 DR  
 XX  
 XX Nucleic acids containing electron-transfer group, useful as labels in  
 PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
 PT a single surface -  
 PT  
 XX  
 XX Example 6; Page 127; 159pp; English.  
 PS  
 XX The present invention relates to a composition comprising two nucleic  
 CC acids each containing an electron-transfer group (ETM) having  
 CC different redox potentials. The invention is used for electronic  
 CC detection of nucleic acids, especially of substitutions (mismatches)  
 CC and single-nucleotide polymorphisms, e.g. for genotyping,  
 CC monitoring gene expression.  
 CC  
 SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match 7.8%; Score 251; DB 22; Length 936;  
 Best Local Similarity 1.0%; Pred. No. 7.7e-31;  
 Matches 8; Conservative 589; Mismatches 184; Indels 0; Gaps 0;  
 QY 1508 gcggtgaaccatcgcacatataatgcttctttaaagcatcttaacatattatt 1567  
 DB 793 GCGGCGCMMWW 734



Db 193 [www.....](#) 134  
Oy 2168 [cagaactccaaatttttttatacgacaagaaataacagatagaacctattt](#) 2227  
Db 133 [www.....](#) 74  
Oy 2228 [gtctggaatgggaagttagtaacatacatcaatgaacaaatttaaataatatagacctat](#) 2287  
Db 73 [www.....](#) 14  
Oy 2288 a 2288  
Db 13 w 13

RESULT 5  
AAf58259/c  
ID AAF58259 standard; DNA; 936 BP.  
XX  
AC AAF58259;  
XX  
DT 24-APR-2001 (first entry)  
XX  
DE Oligonucleotide D2004.  
KW Electron-transfer group; ETM; mismatch; genotyping;  
RV gene expression; ss.  
OS Synthetic.  
XX  
PN WO200107665-A2.  
PD  
XX 01-FEB-2001.  
PE 26-JUL-2000; 2000WO-US20476.  
PR 26-JUL-1999; 99US-0145695.  
PR 17-MAR-2000; 2000US-0190259.  
XX  
PA (CLIN-) CLINICAL MICRO SENSORS INC.  
PI Umek RM;  
FI  
DR WPI; 2001-159728/16.  
XX  
PT Nucleic acids containing electron-transfer group, useful as labels in hybridization assays, e.g. for genotyping, allowing repeat analyses on a single surface -  
PS  
XX Example 6; Page 128; 159pp; English.

The present invention relates to a composition comprising two nucleic acids each containing an electron-transfer group (ETM) having different redox potentials. The invention is used for electronic detection of nucleic acids, especially of substitutions (mismatches) and single-nucleotide polymorphisms, e.g. for genotyping, monitoring gene expression.

Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

	Query Match	7.8%	Score 251;	DB 22;	Length 936;
	Best Local Similarity	1.0%;	Pred. No. 7,7e-31;		
	Matches	8;	Conservative 589;	Mismatches 184;	Indels 0;
				Gaps	0;
Qy	1508	gcgttaaccctcgcgaacttaaatgcttccttaatgatcctttaacatttatt	1567		
	:	:	:	:	:
Db	793	GGGGCGWWWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM	734		
		:	:	:	:
Oy	1568	gttaatgcaatttaataagcgagactgtacaattcaattatatattagatactag	1627		
	: : : :	:	:	:	:
Db	733	wwwmm	674		
		:	:	:	:

[illegible]

RESULT	6
AAAF58262/c	
ID	AAAF58262 standard; DNA; 936 BP.
XX	
AC	AAAF58262;
XX	
DJ	24-APR-2001 (first entry)
XX	
DE	Oligonucleotide D2007.
XX	
KW	Electron-transfer group; ETM; mismatch; genotyping.
KN	gene expression; ss.
XX	
OS	Synthetic.
PN	W0200107665-A2.
PD	
01-FEB-2001.	
XX	
PF	26-JUL-2000; 2000WO-US20476.
PR	26-JUL-1999; 990US-0145695.
PR	17-MAR-2000; 2000US-0190259.

XX (CLIN-) CLINICAL MICRO SENSORS INC.  
 PA Umek RM;  
 XX WPI: 2001-159728/16.  
 DR Nucleic acids containing electron-transfer group, useful as labels in  
 PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
 PT a single surface  
 XX  
 PS Example 6; Page 128; 159pp; English.  
 CC The present invention relates to a composition comprising two nucleic  
 CC acids each containing an electron-transfer group (ETM) having  
 CC different redox potentials. The invention is used for electronic  
 CC detection of nucleic acids, especially of substitutions (mismatches)  
 CC and single-nucleotide polymorphisms, e.g. for genotyping,  
 CC monitoring gene expression.  
 XX  
 SQ Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;

Query Match 7.8%; Score 251; DB 22; Length 936;  
 Best Local Similarity 1.0%; Pred. No. 7.7e-31;  
 Matches 8; Conservative 589; Mismatches 184; Indels 0; Gaps 0;

OY 1508 ggcgttaaccatcgcacataatgcttcttaatgcattcattatatt 1567  
 DB 793 GGGGGCWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 734  
 OY 1568 gttagtgaatttaagaagcgaactgttaacattacattatagatactag 1627  
 DB 733 WWW 674  
 OY 1628 tatgtatattccaatacacttggatgtttaaactatctgttctcctacg 1687  
 DB 673 WWW 614  
 OY 1688 gataataatatacatcgaggttaaaaagtgtctatttcgcgcatgaag 1747  
 DB 613 WWW 554  
 OY 1748 ataaacctaatcttaatttttgaagaatgaaccttaccatagataatcc 1807  
 DB 553 WWW 494  
 OY 1808 gtagtcttggcacaatgacagcctctacaactgtgataagtaattttctgcaa 1867  
 DB 493 WWW 434  
 OY 1868 atattaatagaatctactactacatagagaagaacgcgcatattacatt 1927  
 DB 433 WWW 374  
 OY 1928 aatttaagaacaaattttgaaaaatgtataatttccaacataattataaata 1987  
 DB 373 WWW 314  
 OY 1988 tgcctataatgtattcctatgtcttaaaatatttttatatttagtataaatac 2047  
 DB 313 WWW 254  
 OY 2048 attagaaccataatagttgtgaattccaataatccatataatttttgaatac 2107  
 DB 253 WWW 194  
 OY 2108 caaatatataattagccaataacatgcatagaagttccaaaaaatttggtaa 2167  
 DB 193 WWW 134  
 OY 2168 cagaactccaattttttttataggaagaataacagatagaataatttt 2227  
 DB 133 WWW 74

DB 133 WWW 74  
 OY 2228 gtttggaaatggaatgagcaatacatataagcaattttaaataattataagcct 2287  
 DB 73 WWW 14  
 OY 2288 a 2288  
 DB 13 W 13

## RESULT 7

AAF58255/c  
 ID AAF58255 strand; DNA; 938 BP.

AC AAF58255;

DT 24-APR-2001 (first entry)

DE Oligonucleotide D1876.

KW Electron-transfer group; ETM; mismatch; genotyping;

KW gene expression; ss.

OS Synthetic.

PN W0200107665-A2.

PD 01-FEB-2001.

PF 26-JUL-2000; 2000MO-US20476.

PR 26-JUL-1999; 99US-0145695.

PR 17-MAR-2000; 2000US-0190259.

XX (CLIN-) CLINICAL MICRO SENSORS INC.

PA Umek RM;

DR WPI: 2001-159728/16.

PT Nucleic acids containing electron-transfer group, useful as labels in  
 PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
 PT a single surface  
 XX

PS Example 6; Page 127; 159pp; English.

CC The present invention relates to a composition comprising two nucleic  
 CC acids each containing an electron-transfer group (ETM) having  
 CC different redox potentials. The invention is used for electronic  
 CC detection of nucleic acids, especially of substitutions (mismatches)  
 CC and single-nucleotide polymorphisms, e.g. for genotyping,  
 CC monitoring gene expression.  
 XX

SQ Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;

Query Match 7.8%; Score 251; DB 22; Length 938;  
 Best Local Similarity 1.0%; Pred. No. 7.7e-31;  
 Matches 8; Conservative 589; Mismatches 184; Indels 0; Gaps 0;

OY 1508 ggcgttaaccatcgcacataatgcttcttaatgcattcattatatt 1567  
 DB 793 GGGGGCWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 734  
 OY 1568 gttagtgaatttaagaagcgaactgttaacattacattatagatactag 1627  
 DB 733 WWW 674  
 OY 1628 tatgtatattccaatacacttggatgtttaaactatctgttctcctacg 1687  
 DB 673 WWW 614

```
Oy 1688 gtataaatatcatcagcaggtaaaaaagtttgccttatttcgcgactgacgaag 1747
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 613 wwwmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 554
Oy 1748 ataaacctatgactttaaatttttgaaaatgtaacctttaccatgataatacc 1807
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 553 wwwmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 494
Oy 1808 gtaatttttgcctatgacagcctctaacctgtagtgaatcttttctgcac 1867
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 493 wwwmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 434
Oy 1868 atactaaattaggaattcaatctactacaaatagaagaacagctgagtaattt 1927
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 433 wwwmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 374
Oy 1928 aattcaagacaaaattttgaaaaatgtataatttctacaacataatttaaataga 1987
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 373 wwwmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 314
Oy 1988 tgcctataatgtatctccatgttcttaaatatttttttataattgataataatc 2047
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 313 wwwmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 254
Oy 2048 ataatgaacaaataatgttggtagaatcaaatatctcatataatttttgaactca 2107
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 253 wwwmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 194
Oy 2108 caaatattataattatgtaacaaatgcatagaagaattccaaaaaatttgtaa 2167
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 193 wwwmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 134
Oy 2168 cagaactccaatttttttttattatgacaagaataacagatagaataattt 2227
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 133 wwwmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 74
Oy 2228 gtctggaatggaatgtagtaatacatatgaacaaatttaaaaaattatataagctat 2287
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 73 wwwmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 14
Oy 2288 a 2288
Db 13 w 13

RESULT: 8
AAFS8252 standard; DNA; 936 BP.
AAFS8252;
AC AAF58252;
AT 24-APR-2001 (first entry)
DE Oligonucleotide D1835.
KM Electron-transfer group; ETM; mismatch; genotyping;
KN gene expression; ss.
OS Synthetic.
PN WO200107665-A2.
PD 01-FEB-2001.
PE 26-JUL-2000; 2000WO-US20476.
PR 26-JUL-1999; 99US-0145695.
PA 17-MAR-2000; 2000US-0190259.
PA (CLIN-) CLINICAL MICRO SENSORS INC.
PI umek RM;
XX
```

```
DR WPI; 2001-159728/16.
XX Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PI a single surface.
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 7.8%; Score 250.6; DB 22; Length 936;
Best Local Similarity 0.5%; Pred. No. 8.8e-31;
Matches 4; Conservative 596; Mismatches 185; Indels 0; Gaps 0;

Oy 1544 taatgcatttaacatatattttagtggaaatttaagaagcgaacttgtaacat 1603
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 wwwmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 60
Oy 1604 taacattattatagatagacagatgtagtattccaatacacttggatgtta 1663
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 wwwmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 120
Oy 1664 aacttaactctgtcttctcctacglataaataatcatcaggtaaaaaagtttg 1723
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 121 wwwmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 180
Oy 1724 tcttatttcgcgatgcatgaagataaacctaagttaatttttgaaatgtaac 1783
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 181 wwwmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 240
Oy 1784 cctttactcatagatttaatacgtatgttttgcataatgaacagcctctaacac 1843
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 241 wwwmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 300
Oy 1844 tgtatagtcaaatttttctgcaaatatuaataggaatcactatcaataga 1903
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 301 wwwmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 360
Oy 1904 agaacagctggaattacatttaattaaagacaaaatttgyaaaatgttaatt 1963
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 361 wwwmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 420
Oy 1964 tctaacatattataataatagatgcctataatgtatttctcctagtcttaaatatt 2023
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 421 wwwmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 480
Oy 2024 ttcttataattagttataataatataatgaacaaatagttgtagaattcaaatatc 2083
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 481 wwwmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 540
Oy 2084 tccatttaatttttgaaatctacaaattataatttagtcaataacaatgcataga 2143
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 541 wwwmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 600
Oy 2144 aagttccaaaaaaatttgttaacagaactccaattttttttttagagaacaag 2203
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 601 wwwmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 660
Oy 2204 aaatacagatagaacaaattttagtgaatggaatgtagtaatacatatgaacaa 2263
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 661 wwwmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 720
Oy 2264 tttaaaaaatataatagcctatacgcgctcaaaagtatgtatctagtgatgaat 2323
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 721 wwwmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm 780
```

OY 2324 ataat 2328  
DB 781 www 785

## RESULT 9

AAFS8254  
ID AAF58254 standard; DNA; 936 BP.

AC AAF58254;

DT 24-APR-2001 (first entry)

DE Oligonucleotide D1875.

KW Electron-transfer group; ETM; mismatch; genotyping;  
gene expression; ss.

OS Synthetic.

PN WO200107665-A2.

PD 01-FEB-2001.

PF 26-JUL-2000; 2000WO-US20476.

PR 26-JUL-1999; 99US-0145695.

PR 17-MAR-2000; 2000US-0190259.

PA (CLIN-) CLINICAL MICRO SENSORS INC.

PI Umek RM;

DR WPI: 2001-159728/16.

PT Nucleic acids containing electron-transfer group, useful as labels in  
hybridization assays, e.g. for genotyping, allowing repeat analyses on  
a single surface

PS Example 6; Page 127; 159pp; English.

CC The present invention relates to a composition comprising two nucleic  
acids each containing an electron-transfer group (ETM) having  
different redox potentials. The invention is used for electronic  
detection of nucleic acids, especially of substitutions (mismatches)  
and single-nucleotide polymorphisms, e.g. for genotyping,  
monitoring gene expression.

SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match 7.8%; Score 250.6; DB 22; Length 936;

Best Local Similarity 0.5%; Pred. No. 8.8e-31;

Matches 4; Conservative 596; Mismatches 185; Indels 0; Gaps 0;

OY 1544 taatgacatttaacattttgttgtaatttaataagcgagctgttaact 1603

DB 1 www 60

OY 1604 tacaatattatagactagatgattcaatcaatacacttgatgtta 1663

DB 61 www 120

OY 1664 aacttaacttcttcttctcgaataaataatcaatcgaagtaaaaaagtttg 1723

DB 121 www 180

OY 1724 tcttatttcgagcagcagagataaactaagacttaattttgaaatgttaac 1783

DB 181 www 240

OY 1784 ccttactacatagatttaactcgtatgttttgcacataagacgcctctacaac 1843

DB 241 www 300

OY 1844 tgtatagtaactttttctcgaatataataggaattcaatgactactacaaga 1903

DB 301 www 360

OY 1904 agaaacagctggtatcattttaatttaaaagacaaaattttgaaaaatgttaatt 1963

DB 361 www 420

OY 1964 tctacaatatatttaaaatcagatgcataatgtatttctcactgttcttaaatatt 2023

DB 421 www 480

OY 2024 tttttatatttagtataataacatttgaaacataagttggtgaattcaatalc 2083

DB 481 www 540

OY 2084 tccatttaatttttgaaatctacaattatatttagtcaataacaatgataga 2143

DB 541 www 600

OY 2144 aagttccaaaaaatttgttaacagaaacttccaaatttttttttaigaaacaag 2203

DB 601 www 660

OY 2204 aataacagatagaaactatttgttggaatggaagtagtaataatacattaagcaaa 2263

DB 661 www 720

OY 2264 ttttaaaaaattataagcctacacgcgtcaagtagtatctagtaggtgtaata 2323

DB 721 www 780

OY 2324 ataat 2328

DB 781 www 785

## RESULT 10

AAFS8257  
ID AAF58257 standard; DNA; 936 BP.

AC AAF58257;

DT 24-APR-2001 (first entry)

DE Oligonucleotide D1954.

KW Electron-transfer group; ETM; mismatch; genotyping;  
gene expression; ss.

OS Synthetic.

PN WO200107665-A2.

PD 01-FEB-2001.

PF 26-JUL-2000; 2000WO-US20476.

PR 26-JUL-1999; 99US-0145695.

PR 17-MAR-2000; 2000US-0190259.

PA (CLIN-) CLINICAL MICRO SENSORS INC.

PI Umek RM;

DR WPI: 2001-159728/16.

PT Nucleic acids containing electron-transfer group, useful as labels in  
hybridization assays, e.g. for genotyping, allowing repeat analyses on  
a single surface









[illegible]

Search completed: October 17, 2001, 19:18:28  
Job time: 15908 sec

sequence 44 BP; 19 A; 9 C; 12 G; 10 T; 194 other,

Query Match 2.7%; Score 85.4; DB 22; Length 244;  
Best Local Similarity 3.9%; Pred. No. 3.1e-05;  
Matches 8; Conservative 163; Mismatches 34; Indels 0; Gaps 0



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2001, 14:52:05 ; Search time 156 Seconds

(without alignments)  
3885.732 Million cell updates/sec

Title: US-09-502-426-1\_COPY\_1\_3202

Perfect score: 3202

Sequence: 1 atgtggatatatatgttg.....gagagagagaactagctcc 3202

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PTCUTS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	77.4	2.4	19124	2	US-08-487-826B-13
2	68.6	2.1	837	4	US-08-998-416-288
3	67.4	2.1	19124	2	US-08-487-826B-13
4	65.8	2.1	615	4	US-08-998-416-186
5	65.4	2.0	636	4	US-08-998-416-1137
6	63.6	2.0	8820	2	US-08-446-855A-1
7	63.6	2.0	8920	4	US-09-150-741-1
8	62.6	2.0	4818	3	US-08-817-926-27
9	61.8	1.9	5852	1	US-07-867-106-2
10	60.8	1.9	4818	3	US-08-817-926-27
11	60.6	1.9	837	4	US-08-998-416-288
12	60.4	1.9	8920	2	US-08-446-855A-1
13	60.4	1.9	8920	4	US-09-150-741-1
14	60.2	1.9	665	2	US-08-883-795A-36
15	59.2	1.8	51952	3	US-08-947-823-1
16	59.2	1.8	827	4	US-08-998-416-535
17	58.4	1.8	665	2	US-08-883-795A-36
18	57.8	1.8	1511	1	US-07-991-867B-8
19	57.8	1.8	1511	1	US-08-107-755A-8
20	57.8	1.8	1511	1	US-08-544-332-8
21	57.4	1.8	3095	6	5231168-1
22	56.4	1.8	2058	3	US-08-749-391-1
23	56.4	1.8	2058	3	US-09-390-200-1
24	56.4	1.8	4673	1	US-07-638-431-1
25	56.4	1.8	4673	1	PCT-US92-00018-1
26	56.2	1.8	2960	3	US-08-913-842-3
27	55.6	1.7	701	4	US-08-998-416-701

28	55.6	1.7	5852	1	US-07-867-106-2	Sequence 2, Appl
29	55.2	1.7	1406	3	US-08-913-842-6	Sequence 6, Appl
30	55.2	1.7	1511	1	US-07-991-867B-8	Sequence 8, Appl
31	55.2	1.7	1511	1	US-08-107-755A-8	Sequence 8, Appl
32	55.2	1.7	1511	2	US-08-544-332-8	Sequence 8, Appl
33	55	1.7	1850	3	US-08-617-860B-32	Sequence 32, Appl
34	55	1.7	4098	2	US-08-605-106-4	Sequence 4, Appl
35	54.8	1.7	6768	1	US-08-107-755A-1	Sequence 1, Appl
36	54.8	1.7	8457	1	US-07-991-867B-1	Sequence 1, Appl
37	54.8	1.7	8457	2	US-08-544-332-1	Sequence 1, Appl
38	54.6	1.7	9048	3	US-08-973-273-4	Sequence 1, Appl
39	54.2	1.7	663	4	US-08-998-416-191	Sequence 191, App
40	53.8	1.7	6243	2	US-09-056-075-1	Sequence 1, Appl
41	53.6	1.7	2430	4	US-08-845-258-3	Sequence 3, Appl
42	53.6	1.7	2430	4	US-08-845-258-40	Sequence 40, Appl
43	53.6	1.7	2430	4	US-08-990-571-3	Sequence 3, Appl
44	53.6	1.7	2430	4	US-08-990-571-40	Sequence 40, Appl
45	53.2	1.7	1850	3	US-08-617-860B-32	Sequence 32, Appl

## ALIGNMENTS

RESULT 1  
US-08-487-826B-13  
Sequence 13, Application US/08487826B  
Patent No. 5993827  
GENERAL INFORMATION:  
APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhuan  
APPLICANT: Wellens, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487, 826B  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelson, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH121.001CPI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19124 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTY-SENSE: NO  
US-08-487-826B-13  
Query Match 2.4%; Score 77.4; DB 2; Length 19124;

Best Local Similarity 42.6%; Pred. No. 1.2e-06;  
Matches 590; Conservative 0; Mismatches 781; Indels 14; Gaps 3;

```

OY 1013 ttaaaatacatttcaacgctcactagctgagatctagatagaataagaagacc 1072
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 229 TTATAATACAAATATATATAGTTCCTATTAATAATTAATATATATATATAT 288
OY 1073 gtccaagctattatacaaaagttacaactgaatagcttgaactccttgaaaatt 1132
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 289 TTTCTGTATATTTTAAATATATACAAATTTCTTATTTTAACTTATTCCTTTT 348
OY 1133 tgggaattaccggttctttagtaataatagattagtggtgaacaaataagtaacaaatt 1192
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 349 AATTTCTTAAATTTCTTTATGCAACAAAACATTAAGTAATTCATACATTCACAAAAA 408
OY 1193 agtggcacaacatatacatatcccttaacagaaaacaacttaagaagaagtaacatat 1252
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 409 AAAAAAAAAAAAAAAAAAAAAAAAAATTTATATATATATATATATATATATATATATAT 468
OY 1253 ccataatagtgatagctatctaccccttaacgctatgctataagactaaagaatagttat 1312
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 469 TCAGTATATATATATATATATATATATATATATATATATATATATATATATATATAT 528
OY 1313 gtgagtgctgataaagaattcaacagcggtgtaataattatgggaacgctatgtaagat 1372
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 529 TTTTAACATGAGAAAGATTAGAAATACATTTTCTTTTCTTGAATGCTAATTCACACAT 588
OY 1373 cactgcaaatatcatcttctggttgcacaaataaaacaacaaagaagaagaaga 1432
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 589 ATATATATATATATATATCTTTTAAATTTAAATTAATAATTCCTTTATTTATTTGTT 648
OY 1433 cgaatttcttgcattccatcaatgatcaataatgcatagatcttcttgcgttaacagtt 1492
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 649 TCTTTTATC-----ACATGTGAAATATATTAATAATTTTCGATTTTATTCATATATTT 702
OY 1493 cgaagtcctctcaacgagcggtgtaacacatctgcgaactaataatgcttcttcaatgcatc 1552
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 703 ATGTCGTTATATACCTTAATATAGGCTTTTATACGATATTCATTAATAGAGGTAATAGCCT 762
OY 1553 ttaacaatatatttctggttgcgaatttaacaaagcgaaacttgaacttacaatat 1612
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 763 AATTAATATATATATCTGATTTTATATATATTTCAATATATATTTCCATGAT 822
OY 1613 tatattagatagtagtaigtatctccaatatatacttgcgaatgtttaaactaatc 1672
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 823 TTATTTTCAAAATACATTAATTAATAGATTTCTTAATATTTCTCATTTATTTTATTA 882
OY 1673 ttgttcttccctcaggtataataataatcaatcagaggttaaaaaagtttgcctaat 1732
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 883 GCATATACATGCACATTTTAATAATTAATAAATTTTATTTTAATATATATATACA 942
OY 1733 cgcgaatgcagaagaataaccataatgacttaatttttgaaaaagttaacccttact 1792
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 943 ATTTTCATACATTAATTAATTTTACACACATTTAAGTGTCTCAATATGTAACATTAAT 1002
OY 1793 catagattaaatcacgtaagtgttgcgaataatgcagcctctacaactgtatagt 1852
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1003 AATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1062
OY 1853 caatttttctgcgaataatlaaataaggaatcgaatgcatacacaagaagaagc 1912
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1063 CTGTATTTATCATAGTATATATATATTTGTATTAACGCTTCAAATATATATATATATA 1122
OY 1913 tgaagattacatttaatttaagaacaaatcttgaaaaagttaattcttaacaaat 1972
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1123 TTAATAAATATATATATATATATATTTTGTATGTATATATATATATATATATATAT 1182
OY 1973 attatlaaataatgctcctaataatctcctaattgcttcaataatttttttata 2032
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1183 TAAGATAAAATTCATACCTATTTATTTAAATATATATATATATATATATATATAT 1237
OY 2033 tttagttataataatataatgataagaataaataagtggtgaattcaaatatccattat 2092
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 1238 TATATATATAGTATATAGTATATCAAAATATTTATATATATGTAATTT---ATTAAATTAATAATA 1294
OY 2093 atttttgaactctcaaatatlaatatatttagtcaataacatgcatagaagaattccaa 2152
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1295 TTTGTTATACATACAAAGCTAAAGAAACTATATCAATCTGGATCTAATAGTATATATAT 1354
OY 2153 aaaaaatttgcgttaacagaactccaaatttttttttatayggaacaagaataacag 2212
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1355 ATAAATATCTTTTATATATATATATATGTTCTCTCTTTTATTTTAAATATATATATAT 1414
OY 2213 atagaataactatttgcgttgcgaatggaatgtaataacttaagcaaatlttaaaaaa 2272
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1415 TAATATATTTTTCATATATATATATATATATATATATATATATATATATATATATAT 1474
OY 2273 attataaagcctataacgcgcctcaagaatgctatctagtagtgaatlaataatgcat 2332
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1475 AAAACTTCAAAACATTTTTCATTAATATATATATATATATATATATATATATATATAT 1534
OY 2333 ggtgcgattcagaattggaacacacatgaaacggaatlaaataatlaacttlaaaata 2392
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1535 ACAGAGAAACGTAGAAACATACCAAAAAATTAAGAACAAAAAGAAATATTAACAAAAATTA 1594
OY 2393 ataaa 2397
    |||
Db 1595 TAAAA 1599
    |||

RESULT 2
US-08-998-416-288
; Sequence 288, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jürgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Redischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPIL
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 6239264artis Corporation
; STREET: 3054 Cornwalls Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 288:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 837 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```



```

Db 6387 TATTTTATGATTTTCATGATTTTATGAAATTTTCAAAATTTTATGATATATATATATAA 6328
Qy 1137 attaccggttgatgtaataatagatctagtggttaacaataatgtaataatgta 1196
Db 6327 AATAAGTAAATACACATTTTAAATATATATATATATATATATATATATATATATATAT 6268
Qy 1197 gtaacataacataatccttaacagaaacaaacttaagaaggttaacatcat 1256
Db 6267 TTCATGTTCTATATATTTTATATATATGAAATATTTGTTATATATATATATATATATG 6208
Qy 1257 atatgggtaatgataccttcac-9tatgctaactagaagactaagaatgta 1315
Db 6207 TACTATATATATATATATATATATATATATATATATATATATATATATATATATATAT 6148
Qy 1316 atgctgataatgaatcacaacgctgtaataatataatgagcgtatgtaacatc 1375
Db 6147 ATMAAAATAGTATMAAATCATATATATATATATATATATATATATATATATATATAT 6088
Qy 1376 tgaataatcatctctgctgctgcaacaataaaacaaacaaacaaacaaacaaacaa 1435
Db 6087 TATATATATATATGATATATATATATATATATATATATATATATATATATATATAT 6028
Qy 1436 tttctctggtatcctcaatgataatgataatgataatgataatgataatgataatg 1495
Db 6027 TTTATATTTGTTATATATATATATATATATATATATATATATATATATATATATAT 5968
Qy 1496 agtccctcaacgctgtaacacatcgt--caactataatgctctcttaacatc 1553
Db 5967 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 5908
Qy 1554 ttaacataatctatgtaatgtaatgtaatgtaatgtaatgtaatgtaatgtaatgta 1613
Db 5907 TATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 5848
Qy 1614 atactagatactagatgtaatgtaatgtaatgtaatgtaatgtaatgtaatgta 1673
Db 5847 ACTTATCTATTTTATATATATATATATATATATATATATATATATATATATATAT 5788
Qy 1674 tgtctctcctcagataataataataataataataataataataataataataataata 1733
Db 5787 GATATGATATATATATATATATATATATATATATATATATATATATATATATATAT 5728
Qy 1734 gctgatacgaaggaataacactaactaactaactaactaactaactaactaactaact 1793
Db 5727 TCTATGTA-CATATATATATATATATATATATATATATATATATATATATATATAT 5669
Qy 1794 atagataatcagatgcttctgctcacaataatgagcgcctcacaactgtaagtc 1853
Db 5668 CCAATCATAAACACATATATATATATATATATATATATATATATATATATATATAT 5609
Qy 1854 aattttctcgaatataataataataataataataataataataataataataataata 1913
Db 5608 TATTTGTTGTAATATATATATATATATATATATATATATATATATATATATATATAT 5557
Qy 1914 gctgatacatttaataataataataataataataataataataataataataataata 1973
Db 5558 ---TATTTAAATATTTTCAAAAACTATATATATATATATATATATATATATATAT 5502
Qy 1974 ttattaaataatgagcctcacaataatgctcctccttaataataatcttctatct 2033
Db 5501 TGACATATATATATATATATATATATATATATATATATATATATATATATATATAT 5442
Qy 2034 ttagtataataacataatgagcaataatgctggaatcacaataatcctcacta--- 2090
Db 5441 TTTCTATATATATATATATATATATATATATATATATATATATATATATATATATAT 5382
Qy 2091 ---ataattttgaaatcacaataatataataatgtaataataataatgtaataat 2147
Db 5381 GACTTATATATATATATATATATATATATATATATATATATATATATATATATATAT 5322
Qy 2148 tccaaaaaaatgtgttaacgaaactcacaatcttcttcttaatgtaacaaagaat 2207

```

```

Db 5321 TATTTACTTATTTACTGTTGTAAGATATATATATATATATATATATATATATATAT 5262
Qy 2208 aacagataaataactatctgtgtggaatggaatggaatggaatggaatggaatgga 2267
Db 5261 TAAGCTATTTGGAACATCAAAATCAGATATATATATATATATATATATATATATAT 5202
Qy 2268 aaaaaatataaagcctaacgctcacaatgtaatgtaatgtaatgtaatgtaatgta 2327
Db 5201 AATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 5142
Qy 2328 tgcatagtgctgactcagaat 2348
Db 5141 AAAAAACACAAATCGAAAT 5121

```

```

RESULT 4
US-08-998-416-186
; Sequence 186, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jürgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Redischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSTYPII
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 186:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 615 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: Single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1074RP
; US-08-998-416-186

```

```

Query Match 2.1%; Score 65.8; DB 4; Length 615;
Best Local Similarity 48.3%; Pred. No. 9.9e-05;
Matches 216; Conservative 0; Mismatches 227; Indels 4; Gaps 1;
Qy 1750 aaactaatgacttaatttttgaanaatgtaaccccttactacatagatlaattacgt 1809

```



RESULT 5  
US-08-998-416-1137  
: Sequence 1137 Application US/08998416  
: Patent No. 6239264  
: GENERAL INFORMATION:  
: APPLICANT: Philippesen, Peter  
: APPLICANT: Pohlmann, Rainer  
: APPLICANT: Steiner, Sabine  
: APPLICANT: Mohr, Christine  
: APPLICANT: Wendland, Jurgen  
: APPLICANT: Knechtle, Philipp  
: APPLICANT: Redischung, Corinne  
: TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPPII  
: TITLE OF INVENTION: AND-USES THEREOF  
: NUMBER OF SEQUENCES: 1152  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: No. 6239264artis Corporation  
: STREET: 3054 Cornwallis Road  
: CITY: Research Triangle Park  
: STATE: No. 6239264th Carolina  
: COUNTRY: USA  
: ZIP: 27709  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/998,416  
: FILING DATE: 24-DEC-1997  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: CH 0016/97  
: FILING DATE: 31-DEC-1996  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Meigs, J. Timothy  
: REGISTRATION NUMBER: 38,241  
: REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGCI1976  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 919-541-8687  
: TELEFAX: 919-541-8689

Query Match	2.0%;	Score 65.4;	DB 4;	Length 636;
Best Local Similarity	47.1%;	Pred. No. 0.00012;		
Matches 269;	Conservative 0;	Mismatches 296;	Indels 6;	Gaps 2

OY	1841	aagcgtatgacaaatcttttttcgcgaacatcttbaatctaggaaatcgaatgcacatcatcat	1900
Db	29	AAGATTAAATATAAACCTTTTATTATATATATATTTAAAGTATTTAAATATTTAACTTTT	88
OY	1901	agaagaaacagctgtagtatacatcttbaatcttbaagaacaaaatcttggaaaatgltala	1960
Db	89	ATCATTTATTTATTAATAATTTATTTTGATTTATTAACCTTATTTATTTATTTATTTAA	148
OY	1961	attcctaacaatataatataa-aataagatgcccataatgatacttctcctaagtcctaaat	2019
Db	149	TTTTACTTAATTCACCATTTATTTATTTATTTATTTATTTAAATATTAATTTATTTGAA	208
OY	2020	attcttttctatattagttataataacatatgaacccaataatagttgtagatcca	2079
Db	209	ACTATTTAGTCTATGTCACAAATTTTAAATTTAGTATTATAAATATTTATTTAGATATTTAT	268
OY	2080	latctcaattaaatatttttgaaatcbacaanaattaliaatlltagtcaataacaaatgca	2139
Db	269	TTTTCTTAATTAATTTATTTAAATAGATTTATTCAAATTAATTAATATATTTATTTAAAT	328
OY	2140	tagaagatccaaaaaaatcttgtaaacagaaactccaatctttcttttctatgaa	2199
Db	339	ATTAAA-----ATAATATTTATTTATTTATTTAAAGATTTAAATTTTATTTAAATTTGTAA	383
OY	2200	caagaataacagatagaacaaactatcttgctggaaatggaagttagiatatacatctaa	2258
Db	384	TTATTATTTTATTTATTAATATTCATTTTATTTTAAATATTTATGTTGATTTATTTATTTAA	443
OY	2260	caaatctttaaanaatataataagccatacagcgcccaagatlbgtactagtaggctga	2319
Db	444	CTTTTATTAACAATTTATTTAAATTTAAATTTTAACTTTAAATTTCTTATTTATTTATTTT	503
OY	2320	attaataatgatactgctgcgttcaagaatctgggcaacaatgaaacggaatbaaatt	2379
Db	504	ATAATTTATTTAAATTAATTAATTCATTTTATTTTATTTTATTTATTTATTTAAATTAAT	563
OY	2380	aacttaanaataaataaaatctgagtaaat	2410
Db	564	TAAATTAATATTTTATTCATTTATTTAATTAAT	594
RESULT			
US-08-446-855A-1			
Sequence 1, Application US/08446855A			
Patent No. 5849573			
GENERAL INFORMATION:			
APPLICANT: Stewart, Thomas S			
APPLICANT: Flores, Maria V			
APPLICANT: O'Sullivan, William J			
TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl			
NUMBER OF SEQUENCES: 2			
CORRESPONDENCE ADDRESS:			
ADDRESSEE: Nixon & Vanderhye PC			
STREET: 1100 No. 8649573th Glebe Road, 8th Floor			
CITY: Arlington			
STATE: Virginia			

```

COUNTRY: USA
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,855A
FILING DATE: 06-Jul-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mitchard, Leonard C
REGISTRATION NUMBER: 29, 009
REFERENCE/DOCKET NUMBER: 47-80
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-816-4000
TELEFAX: 703-816-4100
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8920 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic
US-08-446-855A-1

```

```

Query Match      2.0%; Score 63.6; DB 2; Length 8920;
Best Local Similarity 46.8%; Pred. No. 0.00049;
Matches 278; Conservative 0; Mismatches 304; Indels 12; Gaps 2;

QY 1874 aattggaatcaatgctctatacaataagaagaacagctgattatattta 1933
DB 152 aaaaaataataaataaataaataaataaataaataaataaataaataa 211
QY 1934 aagacaataatttgaanaaagtata-attctaacaataatttaaatatgagcct 1992
DB 212 atataatgattatcaataataaataaataaataaataaataaataaataa 271
QY 1993 ataatgatttctctatgctctataaataatttttttatttatttataaataatt 2052
DB 272 ttatatatttatacaataacatttaagttattttattttattttattttatttt 331
QY 2053 gaaccaataatagttggttaattcaataataatcccaataatttttgaacataa 2112
DB 332 tatataataataatagttggttctattttttatttttttttttttttttttt 391
QY 2113 tattaatatttagtcaataaataatgataaagttccaataaataatttgaacaga 2172
DB 392 tattaataatttccaataataataaataaataaataaataaataaataaataa 451
QY 2173 acttccaataattttttttttttttga-----acaagaataacagataagaac 2221
DB 452 ataatgattttacacatttttttttttttttttttttttttttttttttttttt 511
QY 2222 tatttggttggaatggaagtaataatacaataaagaacattttaaataattataa 2281
DB 512 tccccaataatttgggttccctataatttttttttttttttttttttttttttttt 571
QY 2282 gactatacgctcaagaatgtaatactagtagtgtaataataatgcatggtgcatt 2341
DB 572 atttatttttttttttttttttttttttttttttttttttttttttttttttttt 631
QY 2342 cagaatttgggaacaacaatgaaacggaattaaataataactttaataataaataa 2401
DB 632 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 691
QY 2402 tgagtaaatgtgttttctgactatgaggggcaaaaaaagaacatgccaagaag 2455
DB 692 tttaataataatttttttaaacataataataataataataataataataataata 745

```

```

RESULT 7
US-09-150-741-1
Sequence 1, Application US/09150741
Patent No. 6183996
GENERAL INFORMATION:
APPLICANT: Stewart et al.
TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
Patent No. 6183996
TITLE OF INVENTION: Synthetase II
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/09/150,741
CURRENT FILING DATE: 1998-09-10
EARLIER APPLICATION NUMBER: PL6380
EARLIER FILING DATE: 1992-12-16
EARLIER APPLICATION NUMBER: AU93/00617
EARLIER FILING DATE: 1993-12-02
EARLIER APPLICATION NUMBER: 08/446,855
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 8920
TYPE: DNA
ORGANISM: Plasmodium falciparum
US-09-150-741-1

```

```

Query Match      2.0%; Score 63.6; DB 4; Length 8920;
Best Local Similarity 46.8%; Pred. No. 0.00049;
Matches 278; Conservative 0; Mismatches 304; Indels 12; Gaps 2;

QY 1874 aattggaatcaatgctctatacaataagaagaacagctgattatattta 1933
DB 152 aaaaaataataaataaataaataaataaataaataaataaataaataa 211
QY 1934 aagacaataatttgaanaaagtata-attctaacaataatttaaatatgagcct 1992
DB 212 atataatgattatcaataataaataaataaataaataaataaataaataa 271
QY 1993 ataatgatttctctatgctctataaataatttttttatttatttataaataatt 2052
DB 272 ttatatatttatacaataacatttaagttattttataataataataataataata 331
QY 2053 gaaccaataatagttggttaattcaataataatcccaataatttttgaacataa 2112
DB 332 tatataataataatagttggttctattttttatttttttttttttttttttt 391
QY 2113 tattaatatttagtcaataaataatgataaagttccaataaataatttgaacaga 2172
DB 392 tattaataatttccaataataataaataaataaataaataaataaataaataa 451
QY 2173 acttccaataattttttttttttttga-----acaagaataacagataagaac 2221
DB 452 atataatttatacaattcttataataaacaacataataataataataataataa 511
QY 2222 tatttggttggaatggaagtaataatacaataaagaacattttaaataattataa 2281
DB 512 tccccaataatttgggttccctataattttataataatttttttttttttttttt 571
QY 2282 gactatacgctcaagaatgtaatactagtagtgtaataataatgcatggtgcatt 2341
DB 572 atttatttttttttttttttttttttttttttttttttttttttttttttttttt 631
QY 2342 cagaatttgggaacaacaatgaaacggaattaaataataactttaataataaataa 2401
DB 632 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 691
QY 2402 tgagtaaatgtgttttctgactatgaggggcaaaaaaagaacatgccaagaag 2455
DB 692 tttaataataatttttttaaacataataataataataataataataataataata 745

```

RESULT 8

QY	1705	cggagtgaaaaaaggttttgcttatttttcgagtcacgaagaagatcaaccatgacct	1764
Db	4199	ATTATTTATTTTATTACTGTCAAAAGCTTTAAAAAGAGTAGATTATTAACCTTACACTTT	4258
QY	1765	aatttttcgaaaatgtaacccttttact--catagattaatcaacgatgtttttgltg	1821
Db	4259	CTGAGTATATTAATGATATTAACAGTGAATTAATATACCTTTTTTTTCCAAATTTTTTGC	4318
QY	1822	ccataatgacagccctctacaactcgtgatgtaagcaattttctcgcgaataatcaatagga	1881
Db	4319	TTATTCATTTTGGTATTTTGGTATTAATATGCTGAATTAATAAAAAAGCAAAATTTTCCTG	4378
QY	1882	atlocaatgcctactcaataatagaagaacagcgcgaatattacatttaattaaagacaaa	1941
Db	4379	CTTTTGTGTTTATTTCTGTTTAAATTTTTTTTTTGTATCAATCTTTAAATTTAGTTTAA	4438
QY	1942	attttgaaaaatgctataattttctcaacatalttaanaatgatgcctataatgat	2001
Db	4439	ATTTTTTATTTATTTATCTATTTTAAATTTATTAATTTGTTGTAATTTAAATTTATTTAT	4498
QY	2002	ttccatagtctctaaatatatttttttatatttagtataaatacatatgaaccaata	2061
Db	4499	TTCTAAATTTTTTTTGACAAACAAACGTTTTTAAATTTAGAAATTTTAAATTAATTTTAA	4558
QY	2062	atagtgttgatccaataatcccccaatatttttttgaatctcaacattatataat	2121
Db	4559	TTTGTTTTAAATTTGATATACCATTTTTTTTTTTTACTATTTATTCCTATTTTAAATTTTATA	4618
QY	2122	ttagcaaatatacaatgcatalgaagaagttccaaaaaatttltgaacgaaactccaaa	2181
Db	4619	TTATATTCGAAAAAACCTTTAAAAAGACCTGTGTAATTTTATTTGTAATTAACGCAAAA	4678
QY	2182	ttttttttttttatgagacaagaataacaaagatagataaactattgtgtgugaatgaa	2241
Db	4679	TGACATTTATTAAGAAAAAAATTTAAAAAGAAATTAATAAAACAAAGCGTTGATTTTACTTT	4738
QY	2242	gtagtaatacatcaatgaacaaattttaaanaaattatataag	2282
Db	4739	TAACTAAGCTCAAAAGCTCAATATAATTTTCATTTTTTTTTTAAAG	4779

RESULT 9

US-07-667-106-2/c

Sequence 2, Application US/07867106

Patent No. 5389526

GENERAL INFORMATION:

APPLICANT: Slade, Martin B

APPLICANT: Chang, Andy C M

APPLICANT: Williams, Keith L

TITLE OF INVENTION: Improved Plasmid Vectors for Cellular

TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526r1s

STREET: One Liberty Place 46th Floor

CITY: Philadelphia

STATE: PA

COUNTRY: USA

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/867,106

FILING DATE: 19920625

PRIOR APPLICATION DATA:

APPLICATION NUMBER: AU PJ 7187

APPLICATION NUMBER: PCT/AU90/00530

FILING DATE: 02-NOV-1989

ATTORNEY/AGENT INFORMATION:





Query Match 1.9%; Score 60.4; DB 2; Length 8920;  
Best Local Similarity 48.8%; Pred. No. 0.0021;  
Matches 163; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

QY 1919 ttacatttcaatttaagaacaattttcgaanaatgtaataatttctcaacataattt 1978  
DB 8751 TTTTATTTTATTTATTAATTAATTAATTTATTAATTAATTAATTTGTAATTTGAAA 8692  
QY 1979 aaaaatgctgctataatgctatcttccatgcttcttaataattttttttattttagt 2038  
DB 8691 AAACCAATTTGGTTATATACATATGAGTGAATTAATTAATTTGTAATTAATTAATTAAT 8632  
QY 2039 tataaatcatatcatgacaataatagcttggaattcaataatcccaataattt 2098  
DB 8631 TATTAATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8572  
QY 2099 tgaatctcaaatatcatatattagtcataataacatgacataagaagttccaaaaaa 2158  
DB 8571 TTTAAACGTAAATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8512  
QY 2159 ttgtgtaacagaacttcccaatttttttttttgaagaacaagaataacagatagaa 2218  
DB 8511 TATTATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8452  
QY 2219 aactattctgtcgtgaatggaagtagtaata 2252  
DB 8451 AATATACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8418

## RESULT 13

US-09-150-741-1/c  
: Sequence 1, Application US/09150741  
: Patent No. 6183996  
: GENERAL INFORMATION:  
: APPLICANT: Stewart et al.  
: TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate  
: Patent No. 6183996  
: TITLE OF INVENTION: Synthetase II  
: FILE REFERENCE:  
: CURRENT APPLICATION NUMBER: US/09/150,741  
: CURRENT FILING DATE: 1998-09-10  
: EARLIER APPLICATION NUMBER: PL6380  
: EARLIER FILING DATE: 1992-12-16  
: EARLIER APPLICATION NUMBER: AU93/00617  
: EARLIER FILING DATE: 1993-12-02  
: EARLIER APPLICATION NUMBER: 08/446,855  
: EARLIER FILING DATE: 1995-07-06  
: NUMBER OF SEQ ID NOS: 15  
: SOFTWARE: Patent Ver. 2.0  
: SEQ ID NO 1  
: LENGTH: 8920  
: TYPE: DNA  
: ORGANISM: Plasmodium falciparum  
: US-09-150-741-1

Query Match 1.9%; Score 60.4; DB 4; Length 8920;  
Best Local Similarity 48.8%; Pred. No. 0.0021;  
Matches 163; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

QY 1919 ttacatttcaatttaagaacaattttcgaanaatgtaataatttctcaacataattt 1978  
DB 8751 TTTTATTTTATTTATTAATTAATTTATTAATTAATTAATTTATTTGTAATTTGAAA 8692  
QY 1979 aaaaatgctgctataatgctatcttccatgcttcttaataattttttttattttagt 2038  
DB 8691 AAACCAATTTGGTTATATACATATGAGTGAATTAATTAATTTGTAATTAATTAATTAAT 8632  
QY 2039 tataaatcatatcatgacaataatagcttggaattcaataatcccaataattt 2098  
DB 8631 TATTAATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8572

QY 2099 tgaatctcaaatatcatatattagtcataataacatgacataagaagttccaaaaaa 2158  
DB 8571 TTTAAAGCTAAATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8512  
QY 2159 ttgtgtaacagaacttcccaatttttttttttgaagaacaagaataacagatagaa 2218  
DB 8511 TATTATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8452  
QY 2219 aactattctgtcgtgaatggaagtagtaata 2252  
DB 8451 AATATACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8418

## RESULT 14

US-08-883-795A-36  
: Sequence 36, Application US/08883795A  
: Patent No. 5985607  
: GENERAL INFORMATION:  
: APPLICANT: Delcive, Genevieve  
: APPLICANT: Awang, Gregor  
: TITLE OF INVENTION: Recombinant DNA Molecules and Expression  
: TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator  
: NUMBER OF SEQUENCES: 39  
: CORRESPONDENCE ADDRESSES:  
: ADDRESSEE: BERSKIN & PARR  
: STREET: 40 King Street West  
: CITY: Toronto  
: STATE: Ontario  
: COUNTRY: Canada  
: ZIP: M5H 3Y2  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patentin Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/883,795A  
: FILING DATE: 27-JUN-1997  
: CLASSIFICATION: 435  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Gravelle, Micheline  
: REGISTRATION NUMBER: 40,261  
: REFERENCE/DOCKET NUMBER: 7841-062  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (416) 364-7311  
: TELEFAX: (416) 361-1398  
: INFORMATION FOR SEQ ID NO: 36:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 665 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: MOLECULE TYPE: cDNA  
: ORIGINAL SOURCE:  
: ORGANISM: Homo sapiens  
: IMMEDIATE SOURCE:  
: CLONE: Rh 32  
: US-08-883-795A-36

Query Match 1.9%; Score 60.2; DB 2; Length 665;  
Best Local Similarity 46.8%; Pred. No. 0.0013;  
Matches 256; Conservative 0; Mismatches 288; Indels 3; Gaps 2;

QY 1856 ttcttcgcaaatatcaataggaattcactatcatatagaagaacagctcga 1915  
DB 22 TTTTATTAATTAATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 81  
QY 1916 gttatcaatttcaatttaagaacaattttcgaanaatgtaataatttctcaacataatt 1975  
DB 82 TTTTATTAATTAATTAATTTATTAATTAATTTATTAATTAATTAATTAATTAATTAATTAAT 141

```

QY 1976 ataaatatgatcctataatgtaattccctagcttcttaaaatatttttttattt 2035
    || || || || || || || || || || || || || || || || || || || ||
Db 142 TTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 201
    || || || || || || || || || || || || || || || || || || || ||
QY 2036 agttataataactttagaaccataatagttggtgaattcaatccatccatatt 2095
    || || || || || || || || || || || || || || || || || || || ||
Db 202 TTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 261
    || || || || || || || || || || || || || || || || || || || ||
QY 2096 tttagaatctacaattattatt--atttagtcaataacagcatalagaagttccaa 2153
    || || || || || || || || || || || || || || || || || || || ||
Db 262 TTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 321
    || || || || || || || || || || || || || || || || || || || ||
QY 2154 aaaaatttgtaacagaacatcccaatttttttttttttttttttttttttttttt 2213
    || || || || || || || || || || || || || || || || || || || ||
Db 322 TTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 381
    || || || || || || || || || || || || || || || || || || || ||
QY 2214 tagaataactatttggtggtgaatggaagtgaataataacaaatctttaaanaa 2273
    || || || || || || || || || || || || || || || || || || || ||
Db 382 TTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 441
    || || || || || || || || || || || || || || || || || || || ||
QY 2274 ttataaagcctatacgcgcgtcaagtaatgtaattctagtagtggttaataatgcatg 2333
    || || || || || || || || || || || || || || || || || || || ||
Db 442 TTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 500
    || || || || || || || || || || || || || || || || || || || ||
QY 2334 gtagcattcagaattggtgacacaaatgaaacggaattaaatatttaactttaaanaa 2393
    || || || || || || || || || || || || || || || || || || || ||
Db 501 TATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 560
    || || || || || || || || || || || || || || || || || || || ||
QY 2394 taaaaat 2400
    || || || || || || || || || || || || || || || || || || || ||
Db 561 TATAAAT 567
    || || || || || || || || || || || || || || || || || || || ||

```

RESULT 15

```

US-08-947-823-1
; Sequence 1, Application US/08947823
; Patent No. 6114605
; GENERAL INFORMATION:
; APPLICANT: Williamson, Valerie M.
; APPLICANT: Kaloshian, Jasouhi
; APPLICANT: Kaghobli, Jafar
; APPLICANT: Bodeau, John
; APPLICANT: Milligan, Stephen
; TITLE OF INVENTION: Procedures and Materials for Confering
; TITLE OF INVENTION: Pest Resistance in Plants
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/947,823
; FILING DATE: 09-OCT-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/18802
; FILING DATE: 09-OCT-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/028,191
; FILING DATE: 10-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 023070-0702100S
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 51952 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-947-823-1

```

```

Query Match      1.8%; Score 59.2; DB 3; Length 51952;
Best Local Similarity 43.7%; Pred. No. 0.0052;
Matches 410; Conservative 0; Mismatches 518; Indels 10; Gaps 3;

```

```

QY 1524 aactataatgcttctttaaagcatttaacatattatttgtagttagtgaatttaa 1583
    || || || || || || || || || || || || || || || || || || || ||
Db 31954 AACAAATACAAATCAAAATCTTTTAAATGTCATGTATCTCAATGATGACATTTTA 32013
    || || || || || || || || || || || || || || || || || || || ||
QY 1584 taagagcgaactgtaacatataattatttagatctgtagttagtattccaa 1643
    || || || || || || || || || || || || || || || || || || || ||
Db 32014 ATTTAGACGAATTTTCTGTGATATTTTGTGATGTGATCTTCAATTAACAA 32073
    || || || || || || || || || || || || || || || || || || || ||
QY 1644 atacatacttgatggtttaacttaacttcttctcctagtgtaataataatca 1703
    || || || || || || || || || || || || || || || || || || || ||
Db 32074 AAATATCTATCTTTAAATGATTTTATTTATTTAATTAAGGCGAACAATTTATAT 32133
    || || || || || || || || || || || || || || || || || || || ||
QY 1704 tcgaaggttaaaaaagtttgcttatttccgagatgacgaagaataaacttagctt 1763
    || || || || || || || || || || || || || || || || || || || ||
Db 32134 AATTAATGATGAATTTTAAAGAAATTCATACATCTCTGACTTTTAAATTTGACATATA 32193
    || || || || || || || || || || || || || || || || || || || ||
QY 1764 taatttttgaataaagtaaaccttcaatagatatttaacgtaagtttggtgcc 1823
    || || || || || || || || || || || || || || || || || || || ||
Db 32194 AAAGATTTTATTAATTTTAAATGAGAGATTTTATTAATTAATTAATTAATTAATTA 32253
    || || || || || || || || || || || || || || || || || || || ||
QY 1824 ataatacagcctcaaacactgtagatgaattttctcgaataataataatgaat 1883
    || || || || || || || || || || || || || || || || || || || ||
Db 32254 CAATTAACAAAGTGTAGCTGTGCTTGACATGTTTGAAGAAATTTGA--ACTTAT 32311
    || || || || || || || || || || || || || || || || || || || ||
QY 1884 tcaatgctactcaatagaagaagaacagctgagattacatttaataagaacaaat 1943
    || || || || || || || || || || || || || || || || || || || ||
Db 32312 TTTAAATTAATTAATGTTTGAATGATTAATTTTATCTTAATAATTAATTAATTAAT 32371
    || || || || || || || || || || || || || || || || || || || ||
QY 1944 tttagaaatgtaataatcttcaacaatatttaaatatagatgcttaagtatt 2003
    || || || || || || || || || || || || || || || || || || || ||
Db 32372 TGCCTAATCAATGTTTAAATTTGTTTACAGAAATGACCAAAATTCGTAATGATTAAT 32431
    || || || || || || || || || || || || || || || || || || || ||
QY 2004 cctatgcttcaataataatttttttatttagttagtaataataca--ttaagaaccaat 2060
    || || || || || || || || || || || || || || || || || || || ||
Db 32432 TAAATATCTTTAAATTAATGATTAATCAATTTTAAATTAATTAATTAATTAATTAAT 32491
    || || || || || || || || || || || || || || || || || || || ||
QY 2061 aatagttggtgaattcaaatatccatataatatttgaatctacaataattataata 2120
    || || || || || || || || || || || || || || || || || || || ||
Db 32492 AATTAAGCTAAGTTGTTTGAATGATTTTGAAGAAATTAATTAACAAATTTAAATA 32551
    || || || || || || || || || || || || || || || || || || || ||
QY 2121 tttagtcaataacaatgcatagaagaagttccaaaaaaattttagtcaagaacttcaa 2180
    || || || || || || || || || || || || || || || || || || || ||
Db 32552 AAATGTTGTAATGATTAATTTTGTCTTAAACACTTTAAAGAAATTCCTTATGTAA 32611
    || || || || || || || || || || || || || || || || || || || ||
QY 2181 attttttttttaggaagaagaatacaacagatagaagaactatttggtgtaagtcga 2240
    || || || || || || || || || || || || || || || || || || || ||
Db 32612 CATATGTTATATTTGTTAAGATATAGCACACATAATTCGTATCTTATTTGAATATA 32671
    || || || || || || || || || || || || || || || || || || || ||
QY 2241 agtagtaataataacataagaacatttaaaaaattataataagccata-----cgcgctc 2295
    || || || || || || || || || || || || || || || || || || || ||
Db 32672 ATCTTTAAATGACGTATCAACATTTTAAATTTGAAGAAATTAATTAATTAATTAATTA 32731
    || || || || || || || || || || || || || || || || || || || ||
QY 2296 aaagtagtattctagtagagtgtaattataaagcatgagtgagatctagaattgggacaa 2355
    || || || || || || || || || || || || || || || || || || || ||
Db 32732 AAAGAAATTAACAAATGTTTACATTAATTAAGAAATTAATTAATTAATTAATTAATTA 32791
    || || || || || || || || || || || || || || || || || || || ||
QY 2356 caatgaacacgaattaaataatttaacttaaaataaaaaattgagtaagtgtc 2415
    || || || || || || || || || || || || || || || || || || || ||

```

Db 32792 AGATTAAAAATAAAAATCAATTTTAAATAATAATATTTCTTGATATATAATATG 32851  
 QY 2416 ttctgactatlgagggcacaaaaagacatgcacaa 2453  
 Db 32852 ATAACTATTTAACAGGGGAAATTTGACNAGAAAAA 32889

Search completed: October 17, 2001, 19:13:08  
 Job time: 15663 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2001, 14:40:40 ; Search time 5433.22 Seconds  
(without alignments)  
5570.921 Million cell updates/sec

Title: US-09-502-426-1\_COPY\_1\_3202  
Perfect score: 3202

Sequence: 1 atgtgggtatattatgttg;.....gagagagagaactagctcc 3202

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 segs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: gb\_est1:\*
- 2: gb\_est2:\*
- 3: gb\_est3:\*
- 4: gb\_est4:\*
- 5: gb\_est5:\*
- 6: gb\_est6:\*
- 7: gb\_est7:\*
- 8: gb\_est8:\*
- 9: gb\_est9:\*
- 10: gb\_est10:\*
- 11: gb\_est11:\*
- 12: gb\_est12:\*
- 13: gb\_est13:\*
- 14: gb\_est14:\*
- 15: gb\_est15:\*
- 16: gb\_est16:\*
- 17: gb\_est17:\*
- 18: gb\_est18:\*
- 19: gb\_est19:\*
- 20: gb\_est20:\*
- 21: gb\_est21:\*
- 22: gb\_est22:\*
- 23: gb\_est23:\*
- 24: gb\_est24:\*
- 25: gb\_est25:\*
- 26: gb\_est26:\*
- 27: gb\_est27:\*
- 28: gb\_est28:\*
- 29: gb\_est29:\*
- 30: gb\_est30:\*
- 31: gb\_est31:\*
- 32: gb\_est32:\*
- 33: gb\_est33:\*
- 34: gb\_est34:\*
- 35: gb\_est35:\*
- 36: gb\_est36:\*
- 37: gb\_est37:\*
- 38: gb\_est38:\*
- 39: gb\_est39:\*
- 40: gb\_est40:\*
- 41: gb\_est41:\*
- 42: gb\_est42:\*
- 43: gb\_est43:\*

- 44: em\_esthum10:\*
- 45: em\_esthum11:\*
- 46: em\_esthum12:\*
- 47: em\_esthum13:\*
- 48: em\_esthum14:\*
- 49: em\_esthum15:\*
- 50: em\_esthum16:\*
- 51: em\_esthum17:\*
- 52: em\_esthum18:\*
- 53: em\_esthum19:\*
- 54: em\_esthum20:\*
- 55: em\_esthum21:\*
- 56: em\_esthum22:\*
- 57: em\_esthum23:\*
- 58: em\_esthum24:\*
- 59: em\_esthum25:\*
- 60: em\_esthum26:\*
- 61: em\_esthum27:\*
- 62: em\_esthum28:\*
- 63: em\_esthum29:\*
- 64: em\_esthum30:\*
- 65: em\_esthum31:\*
- 66: em\_esthum32:\*
- 67: em\_esthum33:\*
- 68: em\_esthum34:\*
- 69: em\_esthum35:\*
- 70: em\_esthum36:\*
- 71: em\_esthum37:\*
- 72: em\_esthum38:\*
- 73: em\_esthum39:\*
- 74: em\_esthum40:\*
- 75: em\_esthum41:\*
- 76: em\_esthum42:\*
- 77: em\_esthum43:\*
- 78: em\_esthum44:\*
- 79: em\_esthum45:\*
- 80: em\_esthum46:\*
- 81: em\_esthum47:\*
- 82: em\_esthum48:\*
- 83: em\_esthum49:\*
- 84: em\_esthum50:\*
- 85: em\_esthum51:\*
- 86: em\_esthum52:\*
- 87: em\_esthum53:\*
- 88: em\_esthum54:\*
- 89: em\_esthum55:\*
- 90: em\_esthum56:\*
- 91: em\_esthum57:\*
- 92: em\_esthum58:\*
- 93: em\_esthum59:\*
- 94: em\_esthum60:\*
- 95: em\_esthum61:\*
- 96: em\_esthum62:\*
- 97: em\_esthum63:\*
- 98: em\_esthum64:\*
- 99: em\_esthum65:\*
- 100: em\_esthum66:\*
- 101: em\_esthum67:\*
- 102: em\_esthum68:\*
- 103: em\_esthum69:\*
- 104: em\_esthum70:\*
- 105: em\_esthum71:\*
- 106: em\_esthum72:\*
- 107: em\_esthum73:\*
- 108: em\_esthum74:\*
- 109: em\_esthum75:\*
- 110: em\_esthum76:\*
- 111: em\_esthum77:\*
- 112: em\_esthum78:\*
- 113: em\_esthum79:\*
- 114: em\_esthum80:\*
- 115: em\_esthum81:\*
- 116: em\_esthum82:\*

```
117: gb_est148:*
118: gb_est149:*
119: gb_est150:*
120: gb_est151:*
121: gb_est152:*
122: gb_est153:*
123: gb_est154:*
124: gb_est155:*
125: gb_est156:*
126: gb_est157:*
127: gb_est158:*
128: gb_est159:*
129: gb_est160:*
130: gb_est161:*
131: gb_est162:*
132: gb_est163:*
133: gb_est164:*
134: gb_est165:*
135: gb_est166:*
136: gb_est167:*
137: gb_est168:*
138: gb_est169:*
139: gb_est170:*
140: gb_est171:*
141: gb_est172:*
142: gb_est173:*
143: gb_est174:*
144: gb_est175:*
145: gb_est176:*
146: gb_est177:*
147: gb_est178:*
148: gb_est179:*
149: gb_est180:*
150: gb_est181:*
151: gb_est182:*
152: gb_est183:*
153: gb_est184:*
154: gb_est185:*
155: gb_est186:*
156: gb_est187:*
157: gb_est188:*
158: gb_est189:*
159: gb_est190:*
160: gb_est191:*
161: gb_est192:*
162: gb_est193:*
163: gb_est194:*
164: gb_est195:*
165: gb_est196:*
166: gb_est197:*
167: gb_est198:*
168: gb_est199:*
169: gb_est200:*
170: gb_est201:*
171: gb_est202:*
172: gb_est203:*
173: gb_est204:*
174: gb_est205:*
175: gb_est206:*
176: gb_est207:*
177: gb_est208:*
178: gb_est209:*
179: gb_est210:*
180: gb_est211:*
181: gb_est212:*
182: gb_est213:*
183: gb_est214:*
184: gb_est215:*
185: gb_est216:*
186: gb_est217:*
187: gb_est218:*
188: gb_est219:*
189: gb_est220:*

190: gb_est110:*
191: gb_est111:*
192: gb_hrc:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_inv4:*
207: em_gss_inv5:*
208: em_gss_inv6:*
209: em_gss_inv7:*
210: em_gss_inv8:*
211: em_gss_inv9:*
212: em_gss_inv10:*
213: em_gss_inv11:*
214: em_gss_inv12:*
215: em_gss_inv13:*
216: em_gss_inv14:*
217: em_gss_inv15:*
218: em_gss_inv16:*
219: em_gss_inv17:*
220: em_gss_inv18:*
221: em_gss_inv19:*
222: em_gss_inv20:*
223: em_gss_inv21:*
224: em_gss_inv22:*
225: em_gss_inv23:*
226: em_gss_inv24:*
227: em_gss_inv25:*
228: em_gss_inv26:*
229: em_gss_inv27:*
230: em_gss_inv28:*
231: em_gss_inv29:*
232: em_gss_inv30:*
233: em_gss_inv31:*
234: em_gss_inv32:*
235: em_gss_inv33:*
236: em_gss_inv34:*
237: em_gss_inv35:*
238: em_gss_inv36:*
239: em_gss_inv37:*
240: em_gss_inv38:*
241: em_gss_inv39:*
242: em_gss_inv40:*
243: em_gss_inv41:*
244: em_gss_inv42:*
245: em_gss_inv43:*
246: em_gss_inv44:*
247: em_gss_inv45:*
248: em_gss_inv46:*
249: em_gss_inv47:*
250: em_gss_inv48:*
251: em_gss_inv49:*
252: em_gss_inv50:*
253: em_gss_inv51:*
254: em_gss_inv52:*
255: em_gss_inv53:*
256: em_gss_inv54:*
257: em_gss_inv55:*
258: em_gss_inv56:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



Oy	2224	tttcttgtagaaatggaagtagtaacataagaacaattttaaz	2271
Dd	500	TTTTTAAATTAATAATTTTTAAWATTTMwKkkKKkAAWDAGARA	453
RESULT	2		
CNSD0EVL			
LOCUS			
DEFINITION		Drosophila melanogaster genome survey sequence T7 end of BAC:	
		BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.	
ACCESSION		AL069706	
VERSION		GI:4949849	
KEYWORDS		GSS.	
SOURCE		fruit fly.	
ORGANISM		Drosophila melanogaster	
		Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
		Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;	
		Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
REFERENCE		(bases 1 to 1101)	
AUTHORS		Genoscope.	
TITLE		Direct Submission	
JOURNAL		Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage BP 191 91006 Evry Cedex - FRANCE (E-mail : segre@genoscope.cns.fr <a href="#">http://www.genoscope.cns.fr</a> ) Web : <a href="#">www.genoscope.cns.fr</a> ) Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <a href="#">http://www.fruitfly.org/The_BDGP_Drosophila_melanogaster_BAC_library.html</a> . The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mammossier in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2: cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <a href="#">http://bacpac.med.buffalo.edu/drosophila_bac.htm</a> .	
FEATURES			
source		1..1101	
		/organism="Drosophila melanogaster"	
		/db_xref="taxon:7227"	
		/clone_1lb="RPCI-98"	
		/clone="BACR29B23"	
		/note="end : T7"	
BASE COUNT		419 a 91 c 60 g 299 t 232 others	
ORIGIN			
Query Match		3.5%, Score 110.8; DB 219; Length 1101;	
Best Local Similarity		34.5%; Pred. No. 8e-09;	
Matches 223; Conservative 128; Mismatches 290; Indels 5; Gaps 1;			
OY	1513	tcaaccatcgcacactaaatgtcgttcctctaagcatcttaacatatratgttag	1572
Dd	461	TWTMMAMMMNNNAATYTWMMAAAMAATITATNATTAATTAATTAATTAATTTTMMMWMTA	520
OY	1573	ttggaattaagaagggaacctgttaacatacataattaatagataagtagt	1632
Dd	521	TTTWTTTMMWWTTTAAATAAAAAAAAATAATTTAAAMWAATATTAATAATTTMAAWW	580
OY	1633	gatattccaataacatacttgtggagttaaactaatatctgtlctcccaagglata	1692
Dd	561	TATATTAATTAATTAATTAATTTAAATTAATAATTAATTTTMMWTAATAATTAATA	640
OY	1693	aataatacatcgcgaagaaaagatttgccttatcttcgcgcgcgaagataaa	1752
Dd	641	TTTTATTATTATTATTAATAAATAATTTMMWTTATTTAAATTAATTTMAAATAAAAAAAAA	700
OY	1753	cctaagcttaatttttgaaaaatgaaccccttactacatagataaacgtagt	1812

Df	701	AAAAAATWTAANAATAATWTTATWTAATTTTTAAAANAATAAATAAATAAATAATWAAATTAATTWMA	760
Oy	1813	tttctgttcgcaatcagacgcacctcacactgtagtgacaattttctcgcaaat	1872
Df	761	TATATATWTTTWTATMMATWMAWMTATATA-----TAAATATAAAMAAMATATAATATAATMA	815
Oy	1873	aattagggaatccatgcgtcactcatcaatagaagaacagcgtgatcttaattcaatt	1932
Df	816	WATWMAAANAMAMATWMAATWATATWMAATATAMAMAAAATATTAATATWATWATAMMA	875
Oy	1933	aaagcaaaaattttgaaaaatgltataaattctcaacaalatataaatalgagcct	1992
Df	876	AAAAATAMAWMTWTTWTTTWWMAWMATATATAAANAMATAMAAAANAATAAATAAAMA	935
Oy	1993	ataagatcttccccatgctctctaataattttttttaatatattagtataaacacttat	2052
Df	936	WWWTAATATTWTTATTAAMTWATWMAATWTAATWMAWMTWATWMTWATATATATATW	995
Oy	2053	gaaccaataatagttgggtaattccaataatcccataatacttttttgaaactccaat	2112
Df	996	WTAAWTAATATTTTATATATAMWTAATATTTTAAANAAMTATATATATATWMAWMTATWMA	1055
Oy	2113	tattaatattagticaatacaatgcatagaagaagttccaaaaa	2158
Df	1056	AWMAATTATWTTATATATATWTAAMATWMAAAAAMAAMAWMATWMAA	1101
RESULT	3		
CNS00213/c			
LOCUS	CNS00213	1101 bp	DNA
DEFINITION	Drosophila melanogaster genome survey sequence TETJ end of BAC #	BACR05N11 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.	03-JUN-1999
ACCESSION	AL061936		
VERSION	AL061936.1	GI:4940214	
KEYWORDS	GSS:		
SOURCE	ORGANISM	drosophila melanogaster	
		Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
REFERENCE	AUTHORS	Genoscope.	
TITLE	JOURNAL	Direct Submission	
COMMENT		Submitted (03-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)	
		determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's PI and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.	
FEATURES	SOURCE	Location/Qualifiers	
		1..1101	
		/organism="Drosophila melanogaster"	
		/db_xref="taxon:7227"	
		/clone_lib="RPCI-98"	
		/clone="BACR05N11"	
		/note="end : TETJ"	
BASE COUNT	631 a	7 c	28 g
ORIGIN			289 t 146 others



Determination of this BAC-end sequence was carried out as part of collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see [http://www.fruitfly.org/The Drosophila melanogaster BAC library](http://www.fruitfly.org/The%20BDGP/Drosophila%20melanogaster%20BAC%20library.html). was prepared by Kazutoyo Oseegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCR-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

```

FEATURES
    source             1..1101
                        /organism="Drosophila melanogaster"
                        /db_xref="taxon:7227"
                        /clone_1id="RPCI-98"
                        /clone="BACR29P01"
                        /note="end : TET3"

BASE COUNT           366 a      66 c      104 g      351 t      214 others

ORIGIN
Query Match          3.1%; Score 98.2; DB 219; Length 1101;
Best Local Similarity 41.0%; Pred. No. 9.2e-07;
Matches 229; Conservative 77; Mismatches 237; Indels 15; Gaps 2.
```

Oy	1766	tttctccatgatactaattaccgtatggttggcgccatatgaagcctcacaa	cgc	1845
Db	1031	TTTTTMMAMVYACAMMAVTTTATTATTTTMTTMMTMMATVTCWATRTT	TMTAATA	972
Oy	1846	tgatagcaaatcttcctgcacaatltaaataggaaatcaatgcctactaca	tagaag	1905
Db	971	WWAYATTTMWTATACATATATWTTTTTTATATACAAITTTMAAAAATAA	AAMAATACGMNAT	912
Oy	1906	aaagcgcgtagta-ttacattttaattgaagacaaaattctgaaaaagtla	aatt	1964
Db	911	TTTMAAACCATTTTWTTTTAAMWTAAATTTWATTCAGTWTTAWMTAAAA	AAATTTTAAAMT	852
Oy	1965	ctaacaatctatcaaatatgatgcctataatgcatcttcctcgaigtctt	aaataattt	2024
Db	851	AAAAAAAAARWTTTMAATTTTAVATWATRTTWAAMWATATAMATTTTM	WMAATWTTATA	792
Oy	2025	ttttatacttgatcataatacatcatgcaaaccaataagctggtggaatc	ccaatc	2084
Db	791	WNAATTTTWTWTAAMWTTTMAAMWTAATTTTAAMWTTTAAATAAAAAA	ATTTAT	732
Oy	2085	ccatcaatatttttgaagaatccacaatatatacatctatagtcacataa	caagcctagaa	2144
Db	731	TTTTTATTTWTAFTWMAAAATTTTWTTTTMAATTTMWTTTAAATWTTA	AMWTTTAAAMW	672
Oy	2145	agtlccaaaaaaattctgltacaga-----aactccaattttttt		2190
Db	671	WTAATTTTAATTAATTAATTTTAAATTTWMAATATRAAAATTTWMAOWT	RATATWTTT	612
Oy	2191	tttatgtaacaaagaataacagatagaanaactatttgtgtggaatggaag	tagtaata	2250
Db	611	WMCMWTWMAATATATATAATWMAAAATTTTAAATTTWTTTATATWMAA	TAAAAAAMATWMTT	552
Oy	2251	tacattaagcaaatltaaaaaatatataagccctatacgcctcaaagtagt	atctca	2310
Db	551	ATTGTAATTAAMAAATVATTTWMAWTTGAATTTTATTTATTTWTTT	TAAAMTTTATATA	492
Oy	2311	gtaggtgaatcaacaat 2328		
Db	491	TWATTTATTTAATWTTATWT 474		

RESULT 6  
AV7137372/C

LOCUS	AV1717372	807 bp	mRNA	EST	16-Oct-2000
DEFINITION	AV1717372 DCB Homo sapiens cDNA clone DCECH07 5', mRNA sequence.				
ACCESSION	AV1717372				
VERSION	AV1717372.1 GI:10814524				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota: Metazoa; Chordata: Craniata; Vertebrata: Euteleostomi; Mammalia; Eutheria; Primates: Catarrhini; Homnidae; Homo.				
AUTHORS	Xu,X., Gu,J., Liu,F., Qu,J., Zhao,M., Li,Y., Huang,Q., Zhou,J., Song,H., Gu,Y., Yang,Y., Gao,G., Xiao,H., Li,N., Qian,B., Gao,X., Cheng,Z., Xu,S., Gu,W., Tu,Y., Jia,D., Fu,G., Ren,S., Zhong,M., Lu,G., Cheng,Z., and Han,Z.				
TITLE	Homo sapiens cDNA DCB clones				
JOURNAL	unpublished (2000)				
COMMENT	Contact: Zeyuang Han Chinese National Human Genome Center at Shanghai 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China Tel: 86-21-50801919( ext. 45) Fax: 86-21-50801922 Email: hanzgchgc-sh.cn This clone is available at CHGC in Shanghai.				
FEATURES	Location/Qualifiers				
Source	1..807 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="DCECH07" /clone_lib="DCB" /cell_type="dendritic cells" /dev_stage="mature" /lab_host="BM25.8" /note="Vector: pTritEx2; Site_1: sfiIa; Site_2: sfiIb"				
BASE COUNT	388 a 50 c 20 g 345 t 4 others				
ORIGIN					
Query Match	3.1%, Score 98; DB 32; Length 807;				
Best Local Similarity	46.2%; Pred. NO. 1e-06;				
Matches 323; Conservative	0; Mismatches 376; Indels 0; Gaps 0;				
Oy	1577	aatttaataagaagcgaaacttgaacattcaacattatattatagatcagatgatt	1636		
Db	773	AAATTTTGAAATTAATTAATTTATTAATTTAAAGAAATTAATTAATTTAAATTTTAT	714		
Oy	1637	attccaatatacatcattgagatggtttaactaaactctggtcttcctcctcaggtataa	1696		
Db	713	ATTTAATATTTTAAATTAATTTAAATTAATTAATTAATTAATTAATTAATTAATTA	654		
Oy	1697	ttaatcatcgaggtataaaagtttctgctctatcttcgcgagatgcatgaagataaact	1756		
Db	653	AAAAATTAATTAATTAATTAATTTGCTTTATTAATTTATTAATTTAATTAATTAAT	594		
Oy	1757	atgaccttaacttttggaaaatgtaaccttactcatagataaactacgctatgtt	1816		
Db	593	TTATTTTATTTTATTTTATTTAATTAATTAACAATTTTAAATATGATTAATTAAT	534		
Oy	1817	tgttgcataatagcagcctctacaaactgtagatcaattttctgcgaaatlaaat	1876		
Db	533	ATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	474		
Oy	1877	taggaatcacaatgctacatcataatagaagaacagctgagatctatcaatttcaat	1936		
Db	473	TTTGATTTTAGTGGAAATTAATTTAATAAAAAATACAAATGTAATTAATTAATTA	414		
Oy	1937	acaaatcttggaaaatgctataatcttcaacaatcatatcaataatgatgacctaa	1996		
Db	413	ATTAATTAATTAATTAATTAATTTTATGATGATGATTAATTAATTAATTTATTA	354		
Oy	1997	tgatcttcctatgctctaaatatttttttattattatagttatgaataacatcagaac	2056		
Db	353	ATTAAAGTCAAAATTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTAT	294		

Qy 1762 tttaatttttgaaaatgtaaccccttactacataagattaataaccgatglttggtg 1823  
||| |||||| | | ||| |||| : | || ||| |  
Db 1046 ttttttttttaaataataataattttatwtttatatatttatatttttttttttta 987

Query Match      2.98;    Score 94;    DB 219;    Length 1101;





ACCESSION	fly), genomic survey sequence.
VERSION	AL064091
KEYWORDS	AL064091.1 GI:4941847
SOURCE	GSS.
ORGANISM	Insecta; Drosophila melanogaster
AUTHORS	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
TITLE	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
JOURNAL	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 1101)
COMMENT	Drosophila. Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org/The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial ECORI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://Bacpac.med.buffalo.edu/drosophila_bac.htm. Location/Qualifiers 1..1101 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone_lib="RPCI-98" /clone="BACR08K08" /note="end : TET3"
BASE COUNT	395 a 120 c 103 g 334 t 149 others
ORIGIN	
Query Match	2.9%; Score 93.6; DB 219; Length 1101;
Best Local Similarity	42.8%; Pred. No. 5.2e-06;
Matches 205; Conservative 51; Mismatches 218; Indels 5; Gaps 1.	
Oy	1722 tgctatttttcgcatgcataagaagataaacctaatacgacttaatttttgaagaatgta 1781
Db	1062 TTTTAACTTTTTTTTTTTATATATTAATAAATTTTTTTGGTGTGTTTATTTTAAATTTATTTT 1003
Oy	1782 accctttaccatacgaataatccacgtatgcttttgctgcacaatagcacgctctaca 1841
Db	1002 ATTTTATTTTTTTTTWTATTAATAATWAATATAAAAATATTWTTWTAATRTWTTTWARAK 943
Oy	1842 actgtgatagtcgaatttttctgcacaataataataggaaatccaatgcactatcaata 1901
Db	942 AAAAAMAAATMTATTTATTTATTTATTTATTTTATTTTWTATTTATTTATTTAAAAAWTAT 883
Oy	1902 gaagaacaacgcagatattacattcaatttaatttaagaacaaaatttttggaaaaagt----- 1956
Db	882 WAAATATATTTATTTATTTATTTATTTTWTATTTATTTATTTATTAANAATWTTTAAATKAAWTTAAW 823
Oy	1957 tataattctcaacaatatatcaataaaltagcgcataatgatctccatgctcttaa 2016
Db	822 TATMAAATTTTAAAMAAMATATNNNAATATATTAANAATWATATTTTATATTAATAATAA 763
Oy	2017 aataatttttttataattagttacaataacatatgaaccaataagtgvggaatc 2076
Db	762 WTATTTTATTTTAAATATATTTTAAAMAATATATATTTTAAWTTTATTAATAATTTTATTTT 703
Oy	2077 aaatccccaatatatttttgaacataccaaattatataatttatagtgcoataaacat 2136
Db	702 TATTTTATTTTATTTTATTTTATTTTAAATWTTTAAATATAAATAAATTTATTTAAATAATTAAMA 643

OY	2137	gcattgaaagttccaaaaaatlttgttaacagaacttccaatlttttttat	2195
Dd	642	ANATWAAAAAATAAAAAAAAAAAAAAAAAAAAKTWTTTTTTTTTTTTTTT	584
<b>RESULT . 11</b>			
LOCUS	CNSD0E07	1101 bp	DNA
DEFINITION	Drosophila melanogaster genome survey sequence TEM3 end of BAC: BACR29P01 of RpCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.		
ACCESSION	AL069440		
VERSION	ALI069440.1	GI:4949583	GSS.
KEYWORDS	fruit fly.		
SOURCE	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 1101)		
REFERENCE	Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org/The_BDGP_Drosophila_melanogaster_BAC_library.html		
AUTHORS	Aaron Mammosee in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RpCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2: cn bw sp, the same strain used for the BDGP's Pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.		
TITLE	Location/Qualifiers		
JOURNAL	1. 1101 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone_1id="RpCI-98" /clone="BACR29P01" /note="end : TET3"		
<b>FEATURES</b>			
source			
BASE COUNT	366 a	66 c	104 g 351 t 214 others
ORIGIN			
Query Match 2.9%; Score 93.6; DB 219; Length 1101;			
Best Local Similarity 40.4%; Pred. NO. 5.2e-06;			
Matches 246; Conservative 83; Mismatches 269; Indels 11; Gaps 3			
OY	1866	aaatatgaagttaggaatcgaatgtctactcatcaatagaagaacagtcgagtattacatt	1925
Dd	425	ACATGAAMAAAAGCGGAGAACAATTAWTTKCAAGCATATTAATAAAAAARATTAATMAATTA	484
OY	1926	ttaatttaaagcaaaaatttttgaaaatagtataattcttacaatatattataaatt	1985
Dd	485	TAAATWATATTAATAAAMWTTTAAAAAAMAAATTAATAAATAATCAATWTAATTAATTWMT	544
OY	1986	gatgcctaatgtaatttccc-catgtctctaaataatttttttttcatctagtataaa	2044
Dd	545	TATTAATTAAMWATATTTTATTTATMMWATATATAAAAAAATTAATTTTATTTATTTATTT	604
OY	2045	tacattatgaaccaataatagttggatgaattcgaatatcctccattataatttttgaat	2104
Dd	605	WAAGCWMMAAAAAAAMWTTATAAHWTATAATTATATTAATTAATTTTAATTAATTAATTT	664
OY	2105	ctacaatatataatattatgtcgaataaacatgcatagataagttccaaaaaatltyct	2164

[illegible]

RESULT	12
CNS00B01/c	
LOCUS	CNS00B01 1101 bp DNA
DEFINITION	CNS00B01 melanogaster genome survey sequence T7 end of BAC #
	BACR24D09 of RPc1-38 library from <i>Drosophila melanogaster</i> (fruit
	fly), genomic survey sequence.
Accession	AF057410

ACCESSION AL057419 GI:4937885  
VERSION AL057419.1  
KEYWORDS GSS.  
SOURCE fruit fly.  
ORGANISM *Drosophila melanogaster*  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscumorphia; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 1101)

**JOURNAL** Submitted (02-JUN-1999) Genoscope - Centre National de Séquençage : BP 191 91006 Evry cedex - FRANCE (E-mail : segrèf@genoscope.cns.fr - Web : www.genoscope.cns.fr)

**COMMENT** Determination of this BAC-end sequence was carried out as part of a

melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Kammeser in Pieter de Jong's Laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2: cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES	source	location/Qualifiers
1. 1101		
/organism="Drosophila		melanogaster"
/db_xref="taxon:7227"		
/clone_11b="RRC1.98"		
/clone="BACR24D09"		
/note="end : 17"		
420 a	56 c	48 g
		261 t
BASE COUNT		316 others
ORIGIN		

Query Match	2.9%	Score 93;	DB 219;	Length 1101;
Best Local Similarity	39.2%;	Pred. No. 6.5e-06;		
Matches 235;	Conservative 78;	Mismatches 278;	Indels 8;	Gaps 1.

[illegible]

RESULT 13	CNS02Q0V/c	LOCUS	DEFINITION
	CNS02Q0V	661 bp	DNA
			GSS
			14-MAY-2000
			Tetradodon nigroviridis genome survey sequence PUC-Orl end of clone
			158p23 of library G from Tetradodon nigroviridis, genomic survey
			sequence.

ACCESSION AL209800  
 VERSION AL209800.1 GI:7668619  
 KEYWORDS GSS; genome survey sequence.  
 SOURCE Tetraodon nigroviridis.  
 ORGANISM Tetraodon nigroviridis

REFERENCE

ROEST-CROILLIUS, H., Jallou, O., Dasilva, C., Fitzames, C., Fisher, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weisenbach, J.

1 (bases 1 to 661)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorphi; Acanthopterygii; Percormorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.

TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL REFERENCE	unpublished 2 (bases 1 to 661)
AUTHORS	Roest-Crollius, H., Jallou, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fitzames, C., Wincker, P., Brothier, P., Quetier, F., Saulin, W. and Weissenbach, J.
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence



Db 731 RTWRRRRRRRRDMDADADDTARDRRRRRGDGDAGKGGKTKRRRRRRRDRATWDRDTA 672

Qy 1876 ttgagatcaatgctactatacaatagaagaacagctagatattcaatttaattaa 1935  
 Db 671 WWDAAWMTTDTDDMDKRRRRRRRRRTTARAAMDWTWAMDAKMDKTRAD 612

Qy 1936 gacaaaatttggaaaatgataatctcaacaatattataaataatgagctata 1995  
 Db 611 RMDRMAADWTDARKADRWAKARAWRARBRARARADBRWTKGKTATATWTAARA 552

Qy 1996 agtatcttcctatgctctcaaatatttttttataattagttataaataactatga 2055  
 Db 551 AAWMAWMTTATATWMTTWTWTTTWTWTTTAAWMAWMTATWMAWMTTAA 492

Qy 2056 ccaataatgctggtgaatcaaatatccatcaatatttttgaatctcaaatatt 2115  
 Db 491 AAAAAWMTTATATWMTTWTWTTTAAWMTATWMTTWTWTTTAAWMTTAA 432

Qy 2116 taatattgctcaatacagatagaa 2144  
 Db 431 WAATATTTTWTWTTWTAATAAAWMTW 403

RESULT 15

CNS04DOK 945 bp DNA GSS 21-MAY-2000  
 LOCUS Tetraodon nigroviridis genome survey sequence T7 end of clone  
 DEFINITION 101k21 of library G from Tetraodon nigroviridis, genomic survey  
 sequence.  
 AL285149

ACCESSION AL285149.1 GI:8023560  
 VERSION GSS: genome survey sequence.  
 KEYWORDS Tetraodon nigroviridis.  
 SOURCE Tetraodon nigroviridis  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 Acanthomorphi; Acanthopterygii; Perciformes; Tetraodontiformes;  
 Tetraodontidae; Tetraodon.  
 1 (bases 1 to 945)  
 Roest-Crolius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,  
 Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and  
 Weissenbach,J.  
 Characterization and repeat analysis of the compact genome of the  
 freshwater pufferfish Tetraodon nigroviridis

TITLE Unpublished  
 JOURNAL 2 (bases 1 to 945)  
 REFERENCE 2 (bases 1 to 945)  
 AUTHORS Roest-Crolius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,  
 Bernot,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F.,  
 Saurin,W. and Weissenbach,J.  
 Human gene number estimate provided by genome wide analysis using  
 Tetraodon nigroviridis DNA sequence

TITLE Unpublished  
 JOURNAL 3 (bases 1 to 945)  
 REFERENCE 3 (bases 1 to 945)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
 COMMENT This sequence is a single read and was generated as part of a large  
 scale clone-end sequencing project of the Tetraodon nigroviridis  
 genome. For more information, please take a look at  
 http://www.genoscope.cns.fr/Tetraodon.  
 FEATURES  
 source location/Qualifiers  
 1..945  
 /organism="Tetraodon nigroviridis"  
 /db\_xref="taxon:99883"  
 /clone="101H21"  
 /clone\_jib="C"  
 /note="Genoscope sequence ID : C08G101CD11LP1-end : T7"  
 BASE COUNT 386 a 112 c 96 g 231 t 120 others  
 ORIGIN

Matches 248; Conservative 64; Mismatches 276; Indels 11; Gaps 3;

Qy 1816 ttgttgcataatgaagctctacacagctgtagtcaattttctcgaatataa 1875  
 Db 354 TGAAGCCAGCCAGCCAGCCGCSNNBMCTTACCAATCRNNNNNNKTTTNNHMTTAA 413

Qy 1876 ttgagatcaatgctactatacaatagaagaacagctagatattcaatttaattaa 1935  
 Db 414 YANCAFTTAMNCAAMHTTNTTTTTTNTANTAYMTWMTWTTTNTTNTAAT 473

Qy 1936 gacaaaatttggaaaatgataatctcaacaatattataaataatgagctata 1995  
 Db 474 TATATAAATTTAAANWATWMTWMAAATAAATAAATAAATAAATAAATAAATAA 528

Qy 1996 agtatcttcctatgctctcaaatattttttatattagttataaataactatga 2055  
 Db 529 AAAAAATTAATTAATWMTWMTWMTWMTWMTWMTWMTWMTWMTWMTWMTWMT 588

Qy 2056 ccaataatgctggtgaatcaaatatccatcaatatttttgaatctcaaatatt 2115  
 Db 589 ATTTTAAAAAATAATTTTAAATAATAATAATAATAATAATAATAATAATAATA 648

Qy 2116 taatattgctcaatacagatagaaagttccaaaaaatttgcacagaact 2175  
 Db 649 AAWMAWMTTATATWMTTWTWTTTAAWMTATWMTTWTWTTTAAWMTTAA 706

Qy 2176 tccaatttttttttgaagaagaataacagatagaaactatttgcgtgga 2235  
 Db 707 WAAAWMAWMTTATATWMTWMTWMTWMTWMTWMTWMTWMTWMTWMTWMTWMT 766

Qy 2236 atggaagtagtaatacacaatgaagaatttaaaaaataataataacgcgtc 2295  
 Db 767 ATWMTWMTWMTWMTWMTWMTWMTWMTWMTWMTWMTWMTWMTWMTWMTWMT 826

Qy 2296 aaagtagt-----tatctgctggtgtaataatgcatggtgcgcatcgaattg 2351  
 Db 827 AAAAAATTAATTAATWMTWMTWMTWMTWMTWMTWMTWMTWMTWMTWMTWMT 886

Qy 2352 acaacaatgaaacggaatcaaatataacttaaacaaataaataatgcgtaac 2410  
 Db 887 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 945

Search completed: October 17, 2001, 16:25:00  
 Job time: 6260 sec

Query Match 2.8%; Score 90.6; DB 221; Length 945;  
 Best Local Similarity 41.4%; Pred. No. 1.6e-05;

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2001, 19:26:38 ; Search time 9769.47 Seconds  
(without alignments)  
6157.353 Million cell updates/sec

Title: US-09-502-426-1\_COPY\_3000\_6888  
Perfect score: 3889  
Sequence: 1 atgaagctcattggttagtt.....cctagtcagcccaagctt 3889

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_ba3:\*  
4: gb\_in1:\*  
5: gb\_in2:\*  
6: gb\_in3:\*  
7: gb\_on:\*  
8: gb\_ov:\*  
9: gb\_pat1:\*  
10: gb\_pat2:\*  
11: gb\_ph:\*  
12: gb\_p11:\*  
13: gb\_p12:\*  
14: gb\_p13:\*  
15: gb\_p14:\*  
16: gb\_p15:\*  
17: em\_ba1:\*  
18: em\_ba2:\*  
19: em\_fun:\*  
20: em\_htgo\_hum:\*  
21: em\_htgo\_inv:\*  
22: em\_htg\_rod:\*  
23: em\_htg\_hum1:\*  
24: em\_htg\_hum2:\*  
25: em\_htg\_hum3:\*  
26: em\_htg\_hum4:\*  
27: em\_htg\_hum5:\*  
28: em\_htg\_hum6:\*  
29: em\_htg\_hum7:\*  
30: em\_htg\_hum8:\*  
31: em\_htg\_inv1:\*  
32: em\_htg\_inv2:\*  
33: em\_htg\_rod:\*  
34: em\_hum1:\*  
35: em\_hum2:\*  
36: em\_hum3:\*  
37: em\_hum4:\*  
38: em\_hum5:\*  
39: em\_hum6:\*  
40: em\_hum7:\*  
41: em\_in:\*  
42: em\_on:\*  
43: em\_or:\*

44: em\_ov:\*  
45: em\_pat:\*  
46: em\_ph:\*  
47: em\_pl:\*  
48: em\_ro:\*  
49: em\_sts:\*  
50: em\_sy:\*  
51: em\_un:\*  
52: em\_v1:\*  
53: gb\_sts1:\*  
54: gb\_sts2:\*  
55: gb\_sts3:\*  
56: gb\_sy:\*  
57: gb\_un:\*  
58: gb\_v11:\*  
59: gb\_v12:\*  
60: gb\_htg1:\*  
61: gb\_htg2:\*  
62: gb\_htg3:\*  
63: gb\_htg4:\*  
64: gb\_htg5:\*  
65: gb\_htg6:\*  
66: gb\_htg7:\*  
67: gb\_htg8:\*  
68: gb\_htg9:\*  
69: gb\_htg10:\*  
70: gb\_htg11:\*  
71: gb\_htg12:\*  
72: gb\_htg13:\*  
73: gb\_htg14:\*  
74: gb\_htg15:\*  
75: gb\_htg16:\*  
76: gb\_htg17:\*  
77: gb\_htg18:\*  
78: gb\_htg19:\*  
79: gb\_htg20:\*  
80: gb\_htg21:\*  
81: gb\_htg22:\*  
82: gb\_htg23:\*  
83: gb\_htg24:\*  
84: gb\_htg25:\*  
85: gb\_pr1:\*  
86: gb\_pr2:\*  
87: gb\_pr3:\*  
88: gb\_pr4:\*  
89: gb\_pr5:\*  
90: gb\_pr6:\*  
91: gb\_pr7:\*  
92: gb\_pr8:\*  
93: gb\_pr9:\*  
94: gb\_tol:\*  
95: gb\_tol2:\*  
96: gb\_in4:\*  
97: gb\_pr10:\*  
98: em\_ba3:\*

pred. NO. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	3889	100.0	4818	12	AF044216	AF044216 Arabidops
2	3811.4	98.0	84196	14	ATT3A5	AL132979 Arabidops
3	134	3.4	100269	13	ATF18022	AL163817 Arabidops
4	128	3.3	106142	12	AC025417	AC025417 Genomic S
5	94.8	2.4	90459	12	AC012396	AC012396 Arabidops
6	93.2	2.4	87835	12	AB005237	AB005237 Arabidops
7	92.8	2.4	4937	9	AR074364	AR074364 Sequence
8	92.8	2.4	4937	13	ATCYP450D	X87368 A.thaliana



Qy	541	accacagatcgtatcatcgccgtgtagtgcgtggaacttaatagatcatatatacaaaaacgaagaa	600
Db	1470	ACCAACGATCGTATCGATCGATCGTACGACCTGACCTTAATGATTCATATTACAAAACGAAGGAG	1529
Qy	601	gctcttgaaatgtagtataccatacgaagataaagtaggaattctctggaaatgctgaact	660
Db	1530	GCTCTTGAAATGATGATTAATCCATGAAGGTATAGTGGGATCTCTGGGAAATGCTGATGCT	1589
Qy	661	tgtctctgtgtgtagcatgatagatagatagaaatctcgcttaactctcttaatga	720
Db	1590	TGTTCTGTGGTGACATGCATGCATGAGATATGAGAAGGTATCTCGCTTAATCTTAAAGTCA	1649
Qy	721	cgacgcgtcttaagaacctctctacttaaaatgttaagaacactctgttgtctga	780
Db	1650	CGACGCTTTAGAACTATTTCTACTTAAAGATGTGAGACATCACTTGTGTCTTGA	1709
Qy	781	tctctgcaacaaactatacttctctctgcataagaacgaagccaaaagttctatt	840
Db	1710	TTCTTGSCAACAAACCTATTTCTCTGCTCACGACGACCAAAAGTTTTATTTT	1769
Qy	841	tatcttattatctgctaaatttttggttatgaactcttagagttcttaacttttt	900
Db	1770	TATCTTTTATTTTGTCAAAATTTTTTGTATGAATCTTTAAGCTTTTCACTTTT	1829
Qy	901	tcttaacttgacaagtttaagtttaactaataatgacgaagataataatgatactg	960
Db	1830	TTTTAATTTGAACAGTTTACGTTTAATCTAATGCGCAGCATATATGATGATCCTG	1889
Qy	961	gagaagaagaacacagagcaacttaagaagaatgatgaacttcaatgaagaagtgct	1020
Db	1890	GAGAAAGAAAGAAACGAGCAATTTAAAGAAAGATATGTAATCTTCAATGAAGAGCTGTCT	1949
Qy	1021	ctgctctccaaatcaacacgaagcgtcatcaataaaaccttaagaatcatattt	1080
Db	1950	CTGCTCCTCTAATCTACACGAGACGCTTATCATTAAGCTTTAGGTAACTTTATTTT	2009
Qy	1081	tcttctgttaagtcacaaactcactatagattttaaattattattatgtgttaa	1140
Db	2010	TTTTTGCTGTAAGTCAACAACTCATTTATAGGTTTTTAATTTATTTATTTATGCTGTA	2069
Qy	1141	taaaatctcaaaatggttgtgtatgacgaagacaagatgaagttcatcttgagaa	1200
Db	2070	TAAATATCTTAANAATGGTGTGTGATGACGAGCAACGATATTGAAGTTCATTGAGAGAA	2129
Qy	1201	aatggaagaagaaaaattggaatcaagaagaagaagaatgaagaagaagtgaaac	1260
Db	2130	AATGGAAGAGGAAAAATTGATATCAAGGACAGAAGTCAAGAAAGAAAGTGAATAC	2189
Qy	1261	agagagatgaagcagagatgtagatgaagtgatactgtttagaacaacaagaacagacatga	1320
Db	2190	AGAGAGATGAAGCAGAGATGTGATGAGATCATGTATTAGAAACAAACACAGACGATGA	2249
Qy	1321	tctcttggaatggttttgaacatccgaattatcagaagagcaaatctccgatctcat	1380
Db	2250	TCTTTTGGGATGGGTTTGAACAATTCGAAATTTATTCGACGACGACAAATTCGATCTCAT	2309
Qy	1381	tcttagtttgtaatttgcgcgacatgagaactctctctgtagcaattgtctcgatcct	1440
Db	2310	TCTTAGTTTGTATTATGGCCGACATGAGACTCTTCTGTAGCCATTGCTCTGCTATCTT	2369
Qy	1441	ctctctgcaagcttgcctctaaagccgttgaagaagctttagggtaagaatatataacaga	1500
Db	2370	CTTCTTCAACGCTTGCCCTTAACCCCTTGAAGGCTTTAGGGTAAATATTAATAACACA	2429
Qy	1501	caagtttaacttaccacaaattgtctgaatgatataaagtattacagaattatctat	1560
Db	2430	CAAGTTAATTACTACCAAAATGTTAGCTATTAATATAGATTATTAAGAAATTAATCTATA	2489
Qy	1561	gaatatacgaatgaaaaaagtatgatattaaattgtctaacaaatttatgttatgat	1620
Db	2490	GAATATACGATGAAAAAAGATGATATTAATTAATTTGACATTAATTTAAGTTATGATTT	2549

QY	1621	atacttttgaagaagagcatcttgaagtcgagggccgaagaagaactcggagagica	1680
QY	1622	atacttttgaagaagagcatcttgaagtcgagggccgaagaagaactcggagagica	1680
Db	2550	AAATCTTTTGAAGGAAGACCATCTTGTAGATCCGGAGGGCCCAAGAACATAGAGATCA	2609
QY	1681	gaattcaaatcttggaatgatacagaanaatggacttactcaatgctgatactacat	1740
Db	2610	GAATTAATAATGGGATGATTACAGAGAAATGGACTTTACTCAATGCTATGTTACTATCAT	2669
QY	1741	tctcatattatattctatgltcatatglatatglatglatglaacccaatatattgatttttt	1800
Db	2670	TCTCATATTATTATTCATATGTCATATGATTTATTTATGATGAACCAAAATATTGATTTT	2729
QY	1801	tttggtgtgtgtggaaggttataataatgaaactcttcgatctgggaaatgtagtagtttt	1860
Db	2730	TTTTGGTGTGTGGAAGGTATTTAAAGAAACCTTGATTTGGGAAATGTAGTAGGTTTT	2789
QY	1861	gactcgcaagaagcaaccgaagaatgcttcggtacagaaggttaaacttgaatcaaatcttt	1920
Db	2790	GCATGCCAAGACACTCCAAAGATGTCGGTACAAAGSTAAACTTTTACGTACAAAATTTTT	2849
QY	1921	aaaataatgaalccggaatatattgaaactcttatctggatggaanaataatcaaatattca	1980
Db	2850	AAATATATGAATCCGGAAATATTGAATCTTATTTGATGAATAAATTTAAATAATTTACA	2909
QY	1981	ttctctaatgtctggaaaaaagaatacagatactccctagtggtgtggaagtgttaacgggtga	2040
Db	2910	TTTTCTTAAAGTTGGGAAAAAAGATACGATATCCATGTGGGGGAAAGGTTTACCGGTGA	2969
QY	2041	tctcgaagctacatcttggaatacttcggtatgacaacactcaatctcttaactcttgga	2100
Db	2970	TCTCAACCGTACATTTTGGATTTCTCGTTATGACCACCACTTATCTCTTAACTCCTTGA	3029
QY	2101	gatggcaacaggtlaataaanaagttctctcgttcaactatcgaaaaatglatagttt	2160
Db	3030	GATGGCAACAGGTAAATAAAGTTTCTCTCGTTAACTGCAAAATTTGTTATAGTTT	3089
QY	2161	tttcaatctatgcatagaataagaatgaagtcctgaagtgattactactactaagaactatac	2220
Db	3090	TTTTCAATCTATTTGCATGAATAAGATAGCTGCTACGTGATTTTACCTATCTATAGTACTATAC	3149
QY	2221	gagaactatgaatctgggcaaaaacttttactattactaacttcaactgaagtaactctaac	2280
Db	3150	GAGAACTATTATCTGCGCAAAAACTTTTTATTATTTATCTTTCAAGTTAATCTTAAC	3209
QY	2281	acgtcaatggatcatglatcacatlgaaagacataaaatlaaaaaataagaagaagaaga	2340
Db	3210	ACGTCATGATCATTTGATCATACATGAAAGCATTTAAATTTAAATAAGAGAGAAAGAGA	3269
QY	2341	cggtgtgtgtgtgaatgtagtcaagtgaagaacaaatcaatagtgtagtgaatgtatctcttaatgac	2400
Db	3270	CGTGTGTGTGTGAAGTGTAGCTGAMGACATTTAATGTAGTGATGTGTCTTTTAAATGAC	3329
QY	2401	gtgaagagctggcctaaatattctcttaaacgttgaaacgttgatattatattagtaacagagct	2460
Db	3330	GTAGGAGCTGCTTAATATTCTTTATATGCTAATCCGTTATTTATTTATTTATGTCACGGCTT	3389
QY	2461	tgaatacaattgaagatttgaacgagacatgtagtaacaaagcgtcttgaacgattcaacagccc	2520
Db	3390	TGATACAAATTTAAGANTTACGAGAGATGTTACACAGCGCTTTGAGCGATCTCACAGGCC	3449
QY	2521	gatgactgttagctgtcgcttagattctctgccaagtlgaactgtgttttaactagattata	2580
Db	3450	GATGACTTGTAGCTGTAGATTTGAGATTTGCGCACGCTTGACTGTTTAAATCTTGATTTTATA	3509
QY	2581	actctatratattataacaactaactaaatcgggaattgagaatatactatattgatt	2640
Db	3510	ACCTATTTATTTATTAACAATCTATCAAAATCGCGAAATTAAGAGAAATTAATCTATATGTATT	3569
QY	2641	attatgattattatagaataaacttaacttgaataataagaataaactgtgtatcatgatt	2700
Db	3570	ATTATGATTTATTTATGAGTAATACCTTTATGAATAAAGTAATTAATGATGATCATGATGTT	3629
QY	2701	ataagtgtgtggggaaggtlaagaaggtgtgtgaagatgattatgaacccaacgtgtgtgtgt	2760

D	3630	ATAGTGA	TGGGAAG	TAAAG	TGGTGA	GAGATATTA	TAAAG	CAOCC	AGTG	TGGT	3689
Q	2761	gccaac	aagca	cgctgc	cttc	cttc	cttc	cttc	cttc	cttc	2820
D	3690	GCCAA	CAGCA	CGTTC	CTTC	CTTC	CTTC	CTTC	CTTC	CTTC	3749
Q	2821	tgtagt	tataa	aaatcg	gttc	gttc	gttc	gttc	gttc	gttc	2880
D	3750	TGTGAT	TTAT	AAATCG	GTTT	GTTC	TTTT	TTTT	TTTT	TTTT	3809
Q	2881	lcaatc	gcaga	aag	tg	la	gttc	gc	ag	tg	2940
D	3810	TCATC	GTG	CAG	AAATG	GTG	TA	TTTT	TC	AGCT	3869
Q	2941	gggcca	aagc	atgt	gt	tg	tg	tg	tg	tg	3000
D	3870	GGGCCA	AGCGAT	GTG	TG	GTG	TC	AG	CA	TG	3929
Q	3001	catc	tagt	ctt	ct	ta	at	ct	ct	ct	3060
D	3930	CATCA	GTG	TTAA	TTC	CA	TTG	GGA	ATT	AG	3989
Q	3061	tttgt	tgat	ttt	ct	ct	aa	cg	gt	tg	3120
D	3990	TTTTGT	GTAT	TTC	TAA	CGG	TTTG	CC	ATT	AG	4049
Q	3121	aaga	tga	aa	ga	at	tt	ct	ct	ct	3180
D	4050	ANGAT	GAA	AG	AT	TTT	TA	TTC	CT	CT	4109
Q	3181	cccaa	tg	a	t	a	t	a	a	a	3240
D	4110	CCCAAT	GAT	AT	AT	AAAA	TTT	G	AT	AA	4169
Q	3241	gttc	g	a	a	a	a	a	a	a	3300
D	4170	GTTTG	AG	AAAA	GGG	TTT	CG	ACT	TC	CG	4229
Q	3301	tgag	t	ctt	tg	a	c	t	t	g	3360
D	4230	TGAG	CTT	TG	CA	AT	TT	G	AT	TT	4289
Q	3361	ta	a	t	a	t	a	a	a	a	3420
D	4290	TAAAT	G	G	CTT	CT	T	A	A	G	4349
Q	3421	ttc	g	t	t	t	g	t	c	c	3480
D	4350	TTT	G	TTTT	T	G	A	T	T	G	4409
Q	3481	ca	c	t	c	g	t	g	t	c	3540
D	4410	CATC	G	T	G	T	G	T	C	T	4469
Q	3541	acc	t	g	c	a	a	t	a	a	3600
D	4470	ACCT	G	T	CA	AT	T	A	T	A	4529
Q	3601	g	t	g	t	a	c	a	t	a	3660
D	4530	GTTG	T	T	A	AT	T	A	G	A	4589
Q	3661	a	g	g	t	a	a	a	a	a	3720
D	4590	AGG	T	A	A	A	A	A	A	A	4649
Q	3721	a	c	g	t	t	c	t	c	t	3780
D	4650	ACTG	T	T	C	T	T	G	G	C	4709
Q	3781	a	t	a	c	t	t	a	a	a	3840

DB	4770	ATCAGCTTTAGAGATCCTCTCTTCAAGACGATAAGTAATCTTCGTTACCAAGACAATGCT	4769
QY	3841	catctttgaagacaggtctttccaaagtcctagttcaggccaagctt	3889
Db	4770	CATCTTTTGAAAGACAGGCTCTTTTCCAGTCTTCAAGTCTTCAAGGCCAAGCTT	4818
RESULT	2		
ATT3A5/c			
LOCUS	ATT3A5	84196 bp	PLN
DEFINITION	Arabidopsis thaliana	DNA	26-JAN-2000
ACCESSION	AL132979		
VERSION	AL132979.2	GI:6782244	
KEYWORDS			
SOURCE			
ORGANISM			
	thale cress.		
	Arabidopsis thaliana		
	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;		
	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;		
	Brassicales; Brassicaceae; Arabidopsidae.		
REFERENCE	1 (bases 1 to 84196)		
AUTHORS	Bloeker,H., Mewes,H.W., Lemcke,K., Mayer,K.F.X., Quetier,F. and Salanoubat,M.		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 84196)		
AUTHORS	EU Arabidopsis sequencing project.		
TITLE	Direct Submission		
JOURNAL	Submitted (25-JAN-2000) MIPS, at the Max-planck-institut fuer Biochemie, Am Klopferspitz 1ba, D-82152 Martinsried, FRG, E-mail: lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de, project coordinator: Marcel Salanoubat and Francis Quetier, Groupement d'Interet Public, Centre National de Sequencage - GENOSCOPE; 2 rue Gaston Cremieux, BP191, 91006 Evry Cedex, France; http://www.genoscope.cns.fr		
COMMENT	On Jan 27, 2000 this sequence version replaced gi:6424247. Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.		
FEATURES			
source	Location/Qualifiers		
	1..84196		
	/organism="Arabidopsis thaliana"		
	/variety="Columbia"		
	/db_xref="taxon:3702"		
	/chromosome="3"		
LTR	2983..3338		
	/note="355 bp LTR"		
	3339..7564		
	/note="LTR-transposon"		
	7565..7922		
	/note="367bp LTR"		
	12339..13197		
	/gene="T3A5.10"		
	join(12339..12665,12742..12803,12888..13055,13125..13197)		
CDS	/gene="T3A5.10"		
	/note="similarity to cyclin-dependent kinase inhibitor (ICK1), Arabidopsis thaliana, EMBL:AF094772.1"		
	/codon_start=1		
	/product="putative protein"		
	/protein_id="CAB62432.1"		
	/db_xref="GI:6561966"		
	/translation="MAAVRRERDVEENGVTTTYVRRKMEERVDLVESRIISPCVQATRRGIVARNSAGASGVIVRRDSPVEEOCIIEEDSSVSCSTSESKRRLEFDLENNQDRETFSTNYDLNKSSESMNNDSSVADEVESRRIRKLSHETVKEALDEPFOVAEKDILNKLLLECSMKNPDEKDEPLGGGRTEYVKNLP"		
	12339..12665		
exon	/gene="T3A5.10"		
	/number=1		
	12666..12741		
intron	/gene="T3A5.10"		
	/number=1		
	12742..12803		
exon	/gene="T3A5.10"		
	/number=2		
intron	12804..12887		



exon /gene="T3A5.10"  
/number=2  
12888. .13055  
/gene="T3A5.10"  
intron /number=3  
13056. .13124  
/gene="T3A5.10"  
/number=3  
13125. .13197  
/gene="T3A5.10"  
/number=4  
complement(15858. .15901)  
exon /number=1  
15858. .16470  
/gene="T3A5.20"  
complement(join(15858. .15901,16014. .16470))  
/note="similarity to predicted proteins, Arabidopsis  
thaliana"  
CDS /codon\_start=1  
/product="putative protein"  
/protein\_id="CAB62433.1"  
/db\_xref="GI:6561967"  
/translation="MARLEHNSMNGARISFSNEFVIRESEKSNKNNINSRSES  
MPSADFAFVTDYSMIPADEIFLGKILPEKTSVHRTIGEELLTEEGSMVGNTF  
SLRPLSSSSFSFGTWRELLGLKRTHVSRKTDKVNVEVLSQDHKIISGVATREC  
QVADTR"  
intron complement(15902. .16013)  
/number=1  
complement(16014. .16470)  
exon /number=2  
18303. .19931  
/gene="T3A5.30"  
complement(18303. .19931)  
exon /number=1  
complement(18303. .19931)  
/codon\_start=1  
/product="scarecrow-like 7 (SCR7)"  
/protein\_id="CAB62434.1"  
/db\_xref="GI:6561968"  
/translation="MAYWCTDSGNLMAIAOOLIKOKOOOQOHOEOEOEPNPMN  
PSFGTLGSGFSDPQVYNDPGRHFPHLEHONAAVASEEDSDMESLNGSDAS  
QTNDFPIYGHDPFVSPSRLSAPYLNRVKNKDSASQOLPPPPASTAWISPSPPQ  
HPPPPPQDPDLNPIKAIHDYARKPETKPDILIRIKESVSESGD IORGVFAE  
ALSHKETSPSSSSSLEDLILSYKTLNDACPYKFAHLANOAILLATONNNIHI  
VDFGFGIOGSAILOALATRSRSGKPTIRISGIPAPSLGDSPGSLATONRLEDA  
AIIIDNEFYPLPIQLNGSSFRVDPDEVLYVFMLELYKLDLDETTTGTALRLA  
RSLNRIYTLGEVSLNRVEFANRVKNSLRFYSAVFESLEPNLDRDSEKERLVERVL  
FGRRLMDLVRSDDNNKPGTRFGIMEEKEOMRVMEKAGFEPVKPSNVAVSQAKLLM  
NYSSTLSVSESGFISLANNVPLTVSSMR"  
gene 26093. .28999  
/gene="T3A5.40"  
complement(join(26093. .26338,27091. .27200,27308. .27386,  
27478. .27570,27723. .28037,28136. .28288,28374. .28698,  
28779. .28999))  
/note="Contains cytochrome P450 cysteine heme-iron ligand  
signature AA45-464; Prokaryotic membrane lipoprotein lipid  
attachment site AA320-330"  
intron /codon\_start=1  
/product="steroid 22-alpha-hydroxylase (DMF4)"  
/protein\_id="CAB62435.1"  
/db\_xref="GI:6561969"  
/translation="MEFEHHTLPLLLPSILSLFLILKRRNRKTRFLPGKS  
GMPLEGTIGLYKPYATTGLDFMOOHYSKYGKIRSNLFGPTIYVSADAKNRITLO  
NEGRLEECSTYRSIGILGKWSMLVGDHMDMSISUNLSHARKLTILKIDVERH  
TLEVLDSWQNSIESAODEAKFTFNMAKHMSMDPEELETQAKRYVFMKVSZ  
APLNPGTAHYKALOSRATILKFERKMEERLIDKEDEEVEYTKEDAEEMSKSD  
VRKORTDDLLGWLKHSNLSTEOILDLITSLFAGHETSVAIALAIFLOACAKAV  
EELREHLEIARAKKELGESLMDMDYKMMPTOCVIMETLRIGVAVFLHKKALDY  
RYKGDYISGKVLPIYSAVHLDNSRYOPLNFNRWQOONNGNASSSGSGSFTWGN  
NYMPGGGPRILCAGSELAKLEMAVFIHHLVLKFMNLEDDPFAFPVDPENGPIIR  
VSRIL"  
exon complement(26093. .26338)  
/number=1

intron complement(26339. .27090)  
/number=1  
complement(27091. .27200)  
exon /number=2  
complement(27201. .27307)  
intron /number=2  
complement(27308. .27386)  
exon /number=3  
complement(27387. .27477)  
intron /number=3  
complement(27478. .27570)  
exon /number=4  
complement(27571. .27722)  
intron /number=4  
complement(27723. .28037)  
exon /number=5  
complement(28038. .28135)  
intron /number=5  
complement(28136. .28288)  
exon /number=6  
complement(28289. .28373)  
intron /number=6  
complement(28374. .28698)  
exon /number=7  
complement(28699. .28778)  
intron /number=7  
complement(28779. .28999)  
exon /number=8  
38307. .41323  
/gene="T3A5.50"  
complement(38307. .38796)  
/number=1  
complement(join(38307. .38796,38877. .39090,39189. .39270,  
40180. .40242,40337. .40401,40489. .40564,40655. .40734,  
40805. .40919,41225. .41323))  
/note="Contains Eukaryotic putative RNA-binding region  
RNP-1 signature AA179-186"  
/codon\_start=1  
/product="U1 snRNP 70K protein"  
/protein\_id="CAB62436.1"  
/db\_xref="GI:6561970"  
/translation="MGDSGDPLRNPNNAVOARAKVONRANVLOLKLKGSHPTGLTN  
NLKLFEEPRPLEYPPPEKRCPTGTMAQFVSFAPGDEYAPPEVELEPSOKR  
EIRHKLREKGEKAEKLKYPDNDNATGDPYKTLFVSLNSESSESKIKREES  
YGPTRVILVTDQLTNKRGCAFETYNMTRDMKAAVKKADOCKIDGRVILVDVNGRT  
VPMNPRRLGGGLSTRVGGGEIVGEOPGRTSSESPSRPEREKSREKSGERE  
SRRELSSHQPRERSRDREREDKHHNRDROGGDRDRDRDRTDRDGRDRDRDR  
GDRTSRHDHDSRKKEEDYEGGEYEHGGGSRERDAEYRGEPEETRGYYEDOGD  
TDYSHRYDKMEDDEFRYERREYKRSRESREYVR"  
intron complement(38797. .38876)  
/number=1  
complement(38877. .39090)  
exon /number=2  
complement(39091. .39188)  
intron /number=2  
complement(39189. .39270)  
exon /number=3  
complement(39271. .40179)  
intron /number=3  
complement(40180. .40242)  
exon /number=4  
complement(40243. .40336)  
intron /number=4

Query Match 98.0%; Score 3811.4; DB 14; Length 84196;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 3881; Conservative 0; Mismatches 6; Indels 6; Gaps 6;

OY 1 atgaagcattatggtta-gtttaagcattataagaagatttattataatgaagaa 59  
|||||  
DB 29203 ATGAAGCTATTGTTAGGTTAAGCTTAATAAGAGATTATTATAATTAAGACGA 29144  
OY 60 tgataacaattatatttctgactctttaaaaccccccttaacaaacgaagctccct 119





gene	/gene="F18022_10" complement(join(2255..2376,2496..2602,2699..2773, 2896..3004,3094..3217,3419..3493,3582..3743,3871.. 4005,4364..4393,4488..4553,4723..4759,4849..4922,5124.. 5180,5289..5339,5466..5531,5640..5775,5930..5981,6091.. .6118)) /gene="F18022_10" complement(2255..2376) /gene="F18022_10"
exon	/number=1 complement(join(2255..2376,2496..2602,2699..2773, 2896..3004,3094..3217,3419..3493,3582..3743,3871.. 4005,4364..4393,4488..4553,4723..4759,4849..4922,5124.. 5180,5289..5339,5466..5531,5640..5775,5930..5981,6091.. .6118)) /gene="F18022_10" complement(2255..2376) /gene="F18022_10"
CDS	/number=1 complement(join(2255..2376,2496..2602,2699..2773, 2896..3004,3094..3217,3419..3493,3582..3743,3871.. 4005,4364..4393,4488..4553,4723..4759,4849..4922,5124.. 5180,5289..5339,5466..5531,5640..5775,5930..5981,6091.. .6118)) /note="Strong similarity to protoporphyrinogen IX oxidase, glycine max, AB025102" /codon_start=1 /product="protoporphyrinogen oxidase-like protein" /protein_id="CAB87761.1" /db_xref="GI:7573447"
Intron	/translation="MASGAVADHQIEAVSGRRVAVGVSGLAAYKLRSGLNTVT FPADRVRGCKLSVMQGLIMDEGANMTMEAREVEGSLIDDLGREKOQFPISOKRY IYRNGVPVLPTNPRIELVTSVSLOSKFOILLEPFLMKSSKSVSDAEESEVFPE OHFGOEVDYLIDPFVGSTSAAPDLSLMKHSPPDIMNSFSITVGAIRTFEAKGG KRDRKSSPGCTKKGRSGRSFRKGMOILLDPTRCKLSHDETLDKSIVLSLNSGSRQ EWMSLSCVSHNTOIKONPHTDAPLCNVKEKVMKGGOFPOLNLPETINMPLSVLIT ITYTKERVKRPLEGFVLLIPSKBQRKFRTLTGLTFFSMMPPDRSPSDVHLTYTFIGISR NOELAKASTDELQOVVTSDIQLRGVEGPVSNHYMYMKAFPLDVSSDSVWEAIDK MENDLPGFFYAGNHSGGLSVGKSIASCGKAADVLSYLESCNDKKPNDSL" complement(2377..2495) /number=1
exon	complement(2496..2602) /gene="F18022_10" /number=2 complement(2603..2698)
Intron	/number=2 complement(2699..2773) /gene="F18022_10"
exon	/number=3 complement(2774..2895) /number=3 complement(2896..3004) /gene="F18022_10"
Intron	/number=4 complement(3005..3093) /number=4 complement(3094..3217) /gene="F18022_10"
exon	/number=5 complement(3218..3418) /number=5 complement(3419..3493) /gene="F18022_10"
Intron	/number=6 complement(3494..3581) /number=6 complement(3582..3743) /gene="F18022_10"
exon	/number=7 complement(3744..3870) /number=7 complement(3871..4005) /gene="F18022_10"
Intron	/number=8 complement(4006..4363) /number=8 complement(4364..4393) /gene="F18022_10"
exon	/number=9 complement(4394..4487) /number=9 complement(4488..4553) /gene="F18022_10"
Intron	/number=10 complement(4554..4722) /number=10 complement(4723..4759) /gene="F18022_10"
exon	/number=11 complement(4760..4848) /number=11 complement(4849..4922) /gene="F18022_10"
Intron	/number=12 complement(4923..5123) /number=12 complement(5124..5180) /gene="F18022_10"
exon	/number=13 complement(5181..5288) /number=13 complement(5289..5339) /gene="F18022_10"
Intron	/number=14 complement(5340..5465) /number=14 complement(5466..5531) /gene="F18022_10"
exon	/number=15 complement(5532..5639) /number=15 complement(5640..5775) /gene="F18022_10"
Intron	/number=16 complement(5776..5929) /number=16 complement(5930..5981) /gene="F18022_10"
exon	/number=17 complement(5982..6090) /number=17 complement(6091..6118) /gene="F18022_10"
Intron	/number=18 complement(6119..6252) /number=18 complement(6253..6487) /gene="F18022_10"
exon	/number=19 complement(6488..6652) /number=19 complement(6653..6887) /gene="F18022_10"
Intron	/number=20 complement(6888..7122) /number=20 complement(7123..7357) /gene="F18022_10"
exon	/number=21 complement(7358..7592) /number=21 complement(7593..7827) /gene="F18022_10"
Intron	/number=22 complement(7828..8062) /number=22 complement(8063..8297) /gene="F18022_10"
exon	/number=23 complement(8298..8532) /number=23 complement(8533..8767) /gene="F18022_10"
Intron	/number=24 complement(8768..9002) /number=24 complement(9003..9237) /gene="F18022_10"
exon	/number=25 complement(9238..9472) /number=25 complement(9473..9707) /gene="F18022_10"
Intron	/number=26 complement(9708..9942) /number=26 complement(9943..10177) /gene="F18022_10"
exon	/number=27 complement(10178..10412) /number=27 complement(10413..10647) /gene="F18022_10"
Intron	/number=28 complement(10648..10882) /number=28 complement(10883..11117) /gene="F18022_10"
exon	/number=29 complement(11118..11352) /number=29 complement(11353..11587) /gene="F18022_10"
Intron	/number=30 complement(11588..11822) /number=30 complement(11823..12057) /gene="F18022_10"
exon	/number=31 complement(12058..12292) /number=31 complement(12293..12527) /gene="F18022_10"
Intron	/number=32 complement(12528..12762) /number=32 complement(12763..12997) /gene="F18022_10"
exon	/number=33 complement(12998..13232) /number=33 complement(13233..13467) /gene="F18022_10"
Intron	/number=34 complement(13468..13702) /number=34 complement(13703..13937) /gene="F18022_10"
exon	/number=35 complement(13938..14172) /number=35 complement(14173..14407) /gene="F18022_10"
Intron	/number=36 complement(14408..14642) /number=36 complement(14643..14877) /gene="F18022_10"
exon	/number=37 complement(14878..15112) /number=37 complement(15113..15347) /gene="F18022_10"
Intron	/number=38 complement(15348..15582) /number=38 complement(15583..15817) /gene="F18022_10"
exon	/number=39 complement(15818..16052) /number=39 complement(16053..16287) /gene="F18022_10"
Intron	/number=40 complement(16288..16522) /number=40 complement(16523..16757) /gene="F18022_10"
exon	/number=41 complement(16758..16992) /number=41 complement(16993..17227) /gene="F18022_10"
Intron	/number=42 complement(17228..17462) /number=42 complement(17463..17697)<







MRNA	/rpl_family="AT_rich" join(<4436..4769,4878..>4948) /gene="T9L24.29"
gene	4436..4948 /gene="T9L24.29" /note="contains similarity to lipid transfer protein GI:2827141 from (Picea abies)" join(4436..4769,4878..4948) /gene="T9L24.29"
CDS	/codon_start=1 /product="hypothetical protein" /protein_id="AAG30973.1" /db_xref="GI:11120793" /translation="MNIRILGVTVTVILVSVQTFQAFGDVQAMRCVAKIMPCOP YIHISIPPCPCPMKQIAEKDVSCICTAFKHPDLRLFALTRKNAIKILDSGGINH DPSVCKNTNBSPPALPEATSGMHLPTIYN" complement(5193..5216) /rpl_family="AT_rich" complement(5484..5515) /rpl_family="(G)n" join(6312..6594,6690..7310) /gene="T9L24.30"
gene	6312..7310 /gene="T9L24.30" /note="contains similarity to diphosphoinositol polyphosphate phosphohydrolase GI:3978224 from (Homo sapiens)" join(6464..6594,6690..7092) /gene="T9L24.30"
CDS	/codon_start=1 /product="unknown protein" /protein_id="AAG30977.1" /db_xref="GI:11120797" /translation="MNMNIPSKRTIVPPTPESEKVLVSIVSRGDLQRYNTAGRVQV CVPRYKHGGEGIEVLISQKKGMLLPGKGEIDSEIFAALRETFEEAGVQVQ LEESGKNQYKSKRRTMHDGMRPLVSOQFELMPSESEFRQKKVSLSEATLCONS WMREKLEFIRKQCTQ join(<7586..7832,8692..8760,8845..8951,9049..>9171) /gene="T9L24.48"
MRNA	7586..9171 /gene="T9L24.48"
gene	/note="similar to RNA binding glycine rich protein (RGP 1a) GI:436789 from (Nicotiana sylvestris)" join(7386..7832,8692..8760,8845..8951,9049..9171) /gene="T9L24.48"
CDS	/codon_start=1 /product="RNA-binding glycine-rich protein, putative" /protein_id="AAG30979.1" /db_xref="GI:11120799" /translation="MAISLGRVVPSCITSGDRLFIPIRESAICSVSGRINVTGVIS ARRRDIGGVLISSCLSSSPSSSSGPKTKLVYGSFPTEDTLEDEFGN LIHNHWQDKVANKRGFAFLRYTTEEMAKAIQMGKGLDGRVIFVEPAKRRSDMS RAKPRDFPKPOSAPRTFRV" complement(9854..>10321) /gene="T9L24.31"
MRNA	9854..10321 /gene="T9L24.31"
gene	/note="predicted by genemark hmms" /gene="T9L24.31"
CDS	9854..10321 /note="predicted by genemark hmms" /gene="T9L24.31"
repeat_region	/codon_start=1 /product="hypothetical protein" /protein_id="AAG30982.1" /db_xref="GI:11120802"
repeat_region	/translation="MAIEKQDINSSIGFNLISEKRRHNTHDDKPLYVDSLTLNSWA SQPSSTLIIIDDEEVLTEPRKRRKPRGSMWDNVEAVALCCVAKSPPAKADN GVAKASYSFSDTLHETMENKKTINEDKRSLLPEKDXDESTVGSNTKNGMNT" complement(11389..11410) /rpl_family="AT_rich" complement(11574..11604) /rpl_family="AT_rich"
repeat_region	12616..13881 /gene="T9L24.32"

[illegible]



QY 1010 aggaatgtctctcctcctcaatcaccaggaactgtcattacataaagcttcagat 1069  
||| 1 ||||| ||||| 1 1 ||| ||| ||||| 1 |||||  
Db 79025 TGGAGTCTCTCTCTCTCTATTTGACTTGCCTTGTTACATPACAAAGCATGAGCT 78966

QY 1070 acattt 1075  
1 |||  
Db 78965 AAGTTT 78960

RESULT 6  
AB005237/c  
LOCUS AB005237 87835 bp DNA PLN 27-DEC-2000  
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 5, pl clone: MJ3.  
ACCESSION AB005237 BA000015  
VERSION AB005237.1 GI:2264309  
KEYWORDS  
SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone\_lib:Mitsui P1 clone: MJ3.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
REFERENCE  
AUTHORS Sato, S., Kotani, H., Nakamura, Y., Kaneko, T., Asamizu, E., Fukami, M., Miyajima, N. and Tabata, S.  
TITLE Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence features of the 1.6 Mb regions covered by twenty physically assigned P1 clones  
JOURNAL DNA Res. 4 (3), 215-230 (1997)  
MEDLINE 97471969  
REFERENCE 2 (bases 1 to 87835)  
AUTHORS Nakamura, Y.  
TITLE Direct Submission  
JOURNAL Submitted (02-JUL-1997) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: ynakamu@kazusa.or.jp, Tel:81-438-52-3935, Fax:81-438-52-3934)  
COMMENT Address for correspondence: kaos@kazusa.or.jp  
For the latest information on annotation of this clone, please see http://www.kazusa.or.jp/kaos/cgi-bin/agg\_graph.cgi?c=MJ3  
Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'.  
The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, http://compbio.ornl.gov/Grail-1.3/), GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and SplicePredictor (Volker Brendel, Stanford University, http://gremmlin.zoel.iastate.edu/cgi-bin/sp.cgi).  
Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, http://genome.wustl.edu/eddy/tRNAscan-SE/).  
This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions. The 5' clone is M0P10 and the 3' clone is K18J17.  
Location/Qualifiers  
1. 87835  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="5"  
/clone="MJ3"  
/clone\_lib="Mitsui P1"  
join(63, 191,310, .622,711, .1111,1255, .1308,1383, .1487, 1590, .1664,1757, .1801,1946, .2074,2184, .2288,2374, .2442)  
/note="gene\_id: MJ3.1  
sp|P98358"  
/codon\_start=1  
/evidence=not\_experimental  
/product="tubulin gamma-2 chain"

CDS  
/protein\_id="BAB09656.1"  
/db\_xref="GI:9759087"  
/translation="MPREITLYQVGGCNGQIGMEFWKQICLEGISKDGILDEPAGOGDRKDVFFYQADDDQHYIPRALILDEPVIINGTONGERNIRYNENHIFLSHGCGAGNMGASGYHQGKVEEELMDMDREADGSDLEGVLCSHIAGGTSGMGSYLLETILNDKYSKRLVQTVSVFNPQMETSDDVVOYPYNSLTLKRLTLNADCVVLDVLTALNRILVERLHLTNPFAQTSLSTVMSASTTTLRLRYGVNNDLVGLASLIPTRCHFTMTGYTPLTVEROANYIRKTYTLVDVVRRLQTKNIMVSSYANKREASQAKYSILNIIGEDVPPQVHESLQRIREKRLIETIDMGASTQVLSKSPYVQTSRVSGLMASHSITRLFSRCLSQYDKLRKKQATLNDTRKFPMPFRANDLSFEDESRLDIIESLVDEYKACESPDIYKMGMEDPGQLMTGEGNASGVADPKLAF"  
3264, .4736  
/note="contains similarity to cationic amino acid transporter 1"  
gene\_id: MJ3.2"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="BAB09657.1"  
/db\_xref="GI:9759088"  
/translation="MTLSPNLDASQKPRISTENPPPPHISIGVTTGDPATSPAFTVNDIKRTIYLPVLIFFVYSGSGREGIESVKAAGPLAIIVGFIYPPFTMSIPEALITAEKGTLPENGGIVVAVTLAMGPYMGQGWKMLSGVIDNALPILFLDYLSGIPIILSGIIPRAVAILVVALYLYNRGLSTIGVAVALYGLVSLPEVYMSFWSIPIKLPSPRMILVYSKKMGVMSLYLNTLFNMLMDSVSLTGEVENPSKTIIPRALFYALLVVFESYIFPVLTGTGAILDQKLMTQGFADIDKIVGGVGLGMWIOAAATSNMGFLAESMSDFOLGMAERGMLEPFAKRSRRTTPVGLIFSGVITIIISMLSFQIVAAENLYCQGVYLETITPRLRKRYPAASRPRIIPGVGLGSYLMCIIPYVILGIVMAFTNLKVALVSLATVIGLVDQPCLOKQYKGMWLFSTSHLPNME"  
5329, .6300  
/note="gene\_id: MJ3.3  
similar to unknown protein  
sp|P03515"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="BAB09658.1"  
/db\_xref="GI:9759089"  
/translation="MEALSEALSDFDEELVASIADAIAATSPKPLVEFSLHALMAELFKRTKLPKAEVSRFFRAVSIIGLGNIRKQDLERTCGAKGRIIRKVAIIIGTIEIPDKQKRIKVLNRYTFPSRMAIPPHWASQALMSEESFYVRASFPANRPPNQTOKLPPLAHSGFAGLIPRESDDMEFIDLICWCHSAFAEHSLLITPOEOKSPHALVENOKRYTEAARSGFNSDENRIYVLLRGLSSDSLDKMLKVAARVSTLSAQEASVYTRDSFTYVEDRVKIVNDAKMDELKAKRIYDASSSMRSQVAAVDSLGV"  
complement(7255, .7632)  
/note="gb|AA02137.1  
gene\_id: MJ3.4  
similar to unknown protein"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="BAB09659.1"  
/db\_xref="GI:9759090"  
/translation="MDLSEKOIYTHIIPYQTEKSYAACVRRSGMKKPTFFPPHLSARPTQSIKFGDWSGLYDPDIYDGIKISIQGVGVLPRAEITGGVCICLILFLLKILSTYKMSHVSFVSCDL"  
9650, .10282  
/note="gb|AA02134.1  
gene\_id: MJ3.5  
similar to unknown protein"  
/codon\_start=1  
/pseudo  
/evidence=not\_experimental  
complement(join(11836, .11997,12102, .12191,12278, .12653, 12819, .13009,13236, .13403,13534, .13696,14495, .14634, 15024, .16196))  
/note="contains similarity to unknown protein  
gb|AA047871.1  
gene\_id: MJ3.6"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="BAB09660.1"  
/db\_xref="GI:9759091"  
/translation="MTNNAQTATTEFRKSPPOPSQOPISDSDSDSGSSENHQRHNDLSNIFEAIVLDCHSSSSPSIDIAFLSSSSGAVSCILCIERIKRTPPTWSCTSSCFAYHFLFCIQSWARQCLDQAAAVTRPSSNPTEPAVNMNCPGCRSSYQKSKI



OY	742	acttaaaagtgtagagacacactcttgcttgctgtagcttgcgaacaaactcat	801
Db	1402	CATGCTTGATATATATACCGGTTAGTCGCGGTTTAAATTTTAAATCTTGG-----TCTTCTC	1454
OY	802	ttctctgcctaaagcggcgcaaaagtttattttatctcttattttgcctaatt	861
Db	1455	GTTGTTCTCCTCATGGAACAGCAAAAAGGTAACCAAAAAATCTTGCTTATCAAAAAC	1514
OY	862	ttttgtttatgaactcttgagttgacttctaactttttttttaaattgaacagttacgt	921
Db	1515	ATTATATATATATATTTATTTATCGGCTTTCACCTATGTTTTTTTATTAATAAATAAATAAT	1574
OY	922	ttaa	925
Db	1575	AAAA	1578
RESULT	8		
LOCUS	ATCYP450D		
DEFINITION	ATCYP450D	4937 bp	DNA
ACCESSION	A.thaliana	gene cytochrome P450.	PLN
VERSION	X87368.1	GI:871987	07-OCT-1996
KEYWORDS	Cytochrome P450.		
SOURCE	thale cress.		
ORGANISM	Arabidopsis thaliana		
REFERENCE	Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta		
AUTHORS	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II		
TITLE	Brassicales; Brassicaceae; Arabidopsis.		
JOURNAL	1 (bases 1 to 4937)		
MEDLINE	Szekeres,M., Nemeth,K., Koncz-Kalman,Z., Mathur,J., Kauschmann,A.,		
AUTHORS	Altmann,T., Redei,G.P., Nagy,F., Schell,J. and Koncz,C.		
REFERENCE	Brassinosteroids rescue the deficiency of CYP90, a cytochrome P450		
AUTHORS	controlling cell elongation and de-etiolation in Arabidopsis		
JOURNAL	Cell 85 (2), 171-182 (1996)		
TITLE	96200769		
REFERENCE	2 (bases 1 to 4937)		
AUTHORS	Szekeres,M.		
JOURNAL	Direct Submission		
TITLE	Submitted (22-MAY-1995) M. Szekeres, Inst. of Plant Biology,		
REFERENCE	Biological Research Center, Hungarian Academy of Sciences, PO Box		
AUTHORS	521, 6701 Szeged, HUNGARY		
TITLE	3 (bases 1 to 4937)		
JOURNAL	Szekeres,M.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (22-JUN-1995) M. Szekeres, Inst. of Plant Biology,		
TITLE	Biological Research Center, Hungarian Academy of Sciences, PO Box		
JOURNAL	521, 6701 Szeged, HUNGARY		
REFERENCE	4 (bases 1 to 4937)		
AUTHORS	Szekeres,M.		
TITLE	Direct Submission		
JOURNAL	Submitted (26-JUN-1995) M. Szekeres, Inst. of Plant Biology,		
REFERENCE	Biological Research Center, Hungarian Academy of Sciences, PO Box		
AUTHORS	521, 6701 Szeged, HUNGARY		
TITLE	On Jun 28, 1995 this sequence version replaced gi:853720.		
JOURNAL	Location/Qualifiers		
COMMENT	1..4937		
FEATURES	/organism="Arabidopsis thaliana"		
SOURCE	/cultivar="Columbia"		
	/db_xref="taxon:3702"		
	/chromosome="5"		
	/clone_lib="lambda EMBL4"		
	/clone="G206"		
	/map="14.3"		
	755 802		
	join(792..1483,1680..1829,1917..2165,3903..3989,		
	4084..4162,4248..4354,4446..4576,4674..4773)		
	792..1483		
	/number=1		
	join(968..1483,1680..1829,1917..2165,3903..3989,		
	4084..4162,4248..4354,4446..4576,4674..4773)		
	/gene="CYP"		
	join(968..1483,1680..1829,1917..2165,3903..3989,		

Query Match	Best Local Similarity	Matches 225: Conservative	Score 2.4%	Pred. No. 53.1%	DB 13: 0.05	Length 4937	Indels 7	Gaps 1	
4084. 4162.4248.4354.4446.4576.4674.4773)	/gene="CPD"	/function="loss of gene function results in dwarf phenotype"	/note="cytochrome P450"	/product="CYP90 protein"	/protein_id="CAA60794.1"	/db_xref="GI:871988"	/translation="MAPTAFVLLISSTIAAGFLLLRTRRRMGPLPGLIGET FOLIGAKTENPEPIDEVARARYGSVMTLFGEPSTIESADPEPNRPGVGLPEFC STPASIONLKGSHSLMLKSLKRMHSLTMSRANSTIDHMLDIDRLVRENLDSS SSRVLRLMEAAKITEFLTVKQLMSFDGSESRKRYLLVIGFESLPPLFSTYR KAIQARRKVAEALTIVVMKRRREEEBAERKDMALALADGFSDEELVDPLVALL VAGETSTIMTLAKFLETPLALADKEHEKIRAMKSDSYSLSDSVKSMPTQC VNETLTVANIGGVFRAMTDVEIKGYIIPKGMKVFSSRAVLDPNHRKDAETNP MRQNSVTGSPSNVTPPGGPRCLCGYELARVALSVFLRLVGTGSWPAEDKLV FFPTTRQKRPIEVRKRDAT"	1484.1679	1484.1679
exon	/number=1	1680.1829	/gene="CPD"	1830.1916	/number=2	1917.2165	/gene="CPD"	2166.3902	
intron	/number=2	3903.3989	/gene="CPD"	3990.4083	/number=3	4084.4162	/gene="CPD"	4163.4247	
exon	/number=3	4248.4354	/gene="CPD"	4355.4445	/number=4	4446.4576	/gene="CPD"	4577.4673	
intron	/number=4	4674.4915	/number=5	4875.4880	/number=6	4915.5000	/number=7	5000.5000	
exon	/number=5	5000.5000	/number=6	5000.5000	/number=7	5000.5000	/number=8	5000.5000	
intron	/number=6	5000.5000	/number=7	5000.5000	/number=8	5000.5000	/number=9	5000.5000	
exon	/number=7	5000.5000	/number=8	5000.5000	/number=9	5000.5000	/number=10	5000.5000	
intron	/number=8	5000.5000	/number=9	5000.5000	/number=10	5000.5000	/number=11	5000.5000	
exon	/number=9	5000.5000	/number=10	5000.5000	/number=11	5000.5000	/number=12	5000.5000	
intron	/number=10	5000.5000	/number=11	5000.5000	/number=12	5000.5000	/number=13	5000.5000	
exon	/number=11	5000.5000	/number=12	5000.5000	/number=13	5000.5000	/number=14	5000.5000	
intron	/number=12	5000.5000	/number=13	5000.5000	/number=14	5000.5000	/number=15	5000.5000	
exon	/number=13	5000.5000	/number=14	5000.5000	/number=15	5000.5000	/number=16	5000.5000	
intron	/number=14	5000.5000	/number=15	5000.5000	/number=16	5000.5000	/number=17	5000.5000	
exon	/number=15	5000.5000	/number=16	5000.5000	/number=17	5000.5000	/number=18	5000.5000	
intron	/number=16	5000.5000	/number=17	5000.5000	/number=18	5000.5000	/number=19	5000.5000	
exon	/number=17	5000.5000	/number=18	5000.5000	/number=19	5000.5000	/number=20	5000.5000	
intron	/number=18	5000.5000	/number=19	5000.5000	/number=20	5000.5000	/number=21	5000.5000	
exon	/number=19	5000.5000	/number=20	5000.5000	/number=21	5000.5000	/number=22	5000.5000	
intron	/number=20	5000.5000	/number=21	5000.5000	/number=22	5000.5000	/number=23	5000.5000	
exon	/number=21	5000.5000	/number=22	5000.5000	/number=23	5000.5000	/number=24	5000.5000	
intron	/number=22	5000.5000	/number=23	5000.5000	/number=24	5000.5000	/number=25	5000.5000	
exon	/number=23	5000.5000	/number=24	5000.5000	/number=25	5000.5000	/number=26	5000.5000	
intron	/number=24	5000.5000	/number=25	5000.5000	/number=26	5000.5000	/number=27	5000.5000	
exon	/number=25	5000.5000	/number=26	5000.5000	/number=27	5000.5000	/number=28	5000.5000	
intron	/number=26	5000.5000	/number=27	5000.5000	/number=28	5000.5000	/number=29	5000.5000	
exon	/number=27	5000.5000	/number=28	5000.5000	/number=29	5000.5000	/number=30	5000.5000	
intron	/number=28	5000.5000	/number=29	5000.5000	/number=30	5000.5000	/number=31	5000.5000	
exon	/number=29	5000.5000	/number=30	5000.5000	/number=31	5000.5000	/number=32	5000.5000	
intron	/number=30	5000.5000	/number=31	5000.5000	/number=32	5000.5000	/number=33	5000.5000	
exon	/number=31	5000.5000	/number=32	5000.5000	/number=33	5000.5000	/number=34	5000.5000	





Qy 562 tgcctgacttaataagatcattacataatacaaaaggaagagctcttgaatgagatcacc 621  
 Db 302 CCCGGAACGAACCGCTTGTCTTCAGAACGAGGAGCTTTTGTGAGCTTATCC 361  
 Qy 622 tagaagatagtgagatctcttggaagaatgagcagatctgtcttggtgagactga 681  
 Db 362 TCGTTCATTGTTGAACCTTTTGGGAAACACCTCTCTTATGAAAGGTTCTTTGCA 421  
 Qy 682 tagaagatagagaagatctgccttaactcttaagtcacgacgctctagaactctc 741  
 Db 422 TTAACGTATGACACTCTCTACCATGAGCTTTCCTATTTCTTAATCATTTAAAGACCATCT 481  
 Qy 742 acttaagaatgttgagagacataacttgltgtcttgatctctgg 787  
 Db 482 CATGCTTGAATGATGACCGGTTAGTCGGTTTAACTTGTGATCTTGG 527

RESULT 11  
 ATCYP450R 1608 bp mRNA PLN 13-MAY-1996  
 DEFINITION A.thaliana mRNA for cytochrome P450.  
 ACCESSION X87367  
 VERSION X87367.1 GI:853718  
 KEYWORDS Cytochrome P450.  
 SOURCE Chale cress.  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 REFERENCE 1 (bases 1 to 1608)  
 AUTHORS Szekeres,M., Nemeth,K., Koncz-Kalman,Z., Mathur,J., Kauschmann,A., Altman,T., Redei,G.P., Nagy,F., Schell,J. and Koncz,C.  
 TITLE Brassinosteroids rescue the deficiency of CYP90, a cytochrome P450, controlling cell elongation and de-etiolation in Arabidopsis  
 JOURNAL Cell 85 (2), 171-182 (1996)  
 MEDLINE 96200769  
 REFERENCE 2 (bases 1 to 1608)  
 AUTHORS Szekeres,M.  
 TITLE Direct Submission  
 JOURNAL Submitted (22-MAY-1995) M. Szekeres, Inst. of Plant Biology, Biological Research Center, Hungarian Academy of Sciences, PO Box 521, 6701 Szeged, HUNGARY

FEATURES  
 source  
 1..1608  
 /organism="Arabidopsis thaliana"  
 /cultivar="Columbia"  
 /db\_xref="taxon:3702"  
 /chromosome="5"  
 /clone\_lib="lambda gt10"  
 /clone="C204"  
 /map="14.3 cm"  
 48..1466  
 /gene="CYP90"  
 /gene="CPD"  
 /function="loss of gene function results in dwarf phenotype"  
 /note="cytochrome P450"  
 /codon\_start=1  
 /product="CYP90 protein"  
 /protein\_id="CAA60793.1"  
 /db\_xref="GI:853719"  
 /db\_xref="SWISS-PROT:Q42569"  
 /translation="MAFTAFLLLSIAAGFLILRLRRYRMKGLPGSGILPTIGF  
 FOLGAYKTEPPEPFDERVARYSVPTHTLFGPTTFSADPTENRYLQNEGLFEC  
 STPASICLGLKSLHLKMSLHRLMSLTWSPANSSITIKDMLADIDRLRMLDSW  
 SSRVLMEARKITFELTVKQMSFDPGEMSESLRKEYLIVTEGFSLPLTEFTYR  
 KAIGARRVAALIVVVMKRRREEEGEERKMDLAALLADGFSDEIYDFVAL  
 VAGETSTIMTLAVKELTETPLALQKEHEKIRAMKSDSYSLSDSYKSMPTCC  
 VVNETLRANITIGVPRAMTDVEIKGIKPKGMKVSPPFAVLDDNHFEDATFNP  
 WRMSNSVTTPSPSVTFPPGGPRLCGEYLARALSYFLHRLVLTGFSWPAEDKLY  
 FFPTTRDKRPIPIVKKRDEAT"

BASE COUNT 382 a 374 c 374 g 478 t

## ORIGIN

Query Match 2 3% Score 87.6; DB 13; Length 1608;  
 Best Local Similarity 56.6%; Pred. No. 0.00012;  
 Matches 162; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

Qy 502 cagcatgttagagataatagatcgaacttggttgagagacagatcgtacagctga 561  
 Db 242 CCGGTACGGTTGGGTTTTCATGACCATCTTTTGTGTAACGACGATTTTCTACGCTCA 301  
 Qy 562 tgcctgacttaataagatcattacataatacaaaaggaagagctcttgaatgagatcacc 621  
 Db 302 CCCGGAACGAACCGCTTGTCTTCAGAACGAGGAGCTTTTGTGAGCTTATCC 361  
 Qy 622 tagaagatagtgagatctcttggaagaatgagcagatctgtcttggtgagactga 681  
 Db 362 TCGTTCATTGTTGAACCTTTTGGGAAACACCTCTCTTATGAAAGGTTCTTTGCA 421  
 Qy 682 tagaagatagagaagatctgccttaactcttaagtcacgacgctctagaactctc 741  
 Db 422 TTAACGTATGACACTCTCTACCATGAGCTTTCCTATTTCTTAATCATTTAAAGACCATCT 481  
 Qy 742 acttaagaatgttgagagacataacttgltgtcttgatctctgg 787  
 Db 482 CATGCTTGAATGATGACCGGTTAGTCGGTTTAACTTGTGATCTTGG 527

RESULT 12  
 AB009048/C  
 LOCUS AB009048 71736 bp DNA PLN 27-DEC-2000  
 DEFINITION Arabidopsis thaliana genomic DNA, chromosome 5, TAC clone:K15B6.  
 ACCESSION AB009048 BA000015  
 VERSION AB009048.1 GI:2656024  
 KEYWORDS  
 SOURCE  
 ORGANISM Arabidopsis thaliana (strain:Columbia) DNA, clone\_lib:MLtsul TAC clone:K15B6.  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 REFERENCE 1 (sites)  
 AUTHORS Sato,S., Kaneko,T., Kotani,H., Nakamura,Y., Asamizu,E., Miyajima,N. and Tabata,S.  
 TITLE Structural analysis of Arabidopsis thaliana chromosome 5. IV. Sequence features of the regions of 1,456,315 bp covered by nineteen physically assigned P1 and TAC clones  
 JOURNAL DNA Res. 5 (1), 41-54 (1998)  
 MEDLINE 98290546  
 REFERENCE 2 (bases 1 to 71736)  
 AUTHORS Nakamura,Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (27-NOV-1997) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp, Tel:81-438-52-3935, Fax:81-438-52-3934)  
 COMMENT Address for correspondence: kaos@kazusa.or.jp  
 For the latest information on annotation of this clone, please see [http://www.kazusa.or.jp/kaos/cg1-bin/agg\\_graph.cgi?c=K15B6](http://www.kazusa.or.jp/kaos/cg1-bin/agg_graph.cgi?c=K15B6)  
 Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'.  
 The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, <http://compbio.ornl.gov/Grail-1.3/>), GENSCAN (Chris Burge, MIT, <http://CCR-081.mit.edu/GENSCAN.html>), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, <http://www.cbs.dtu.dk/services/NetGene2/>) and SplicePredictor (Volker Brendel, Stanford University, <http://agrimini1.zool.iastate.edu/cg1-bin/sp.cgi>).  
 Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, <http://genome.wustl.edu/eddy/tRNAscan-se/>).



unknown protein"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="BAB08644.1"  
/db\_xref="gi:9758065"  
/translation="MQLSSSSAPPEALILEMLOKEMGYRQLGPNYSSKSHVPSIDAI  
RKIRGNMIPWELINFEKSEKIVERRNITVHGSSNSISVSPGSESKG  
RKDKTVGESSVAEDREALOERLAKVERLRNIVRRQDKLAKMLEVREEA  
ERKMLDERANVRHQLLEAYDOOCDEATRFAYEYHRLQVYNOADQSVSSN  
EVLSSANSREAVSYVKGCTKSADVDIIMETTERNIRIYCDLLASRMIERNSF  
PAYENGICISPELETFATLGEFYEGEITDEKVTIVNSLRPPLLQIAIAYTLRIKT

Query Match 2.0%; Score 79.4; DB 12; Length 71736;  
Best Local Similarity 49.7%; Pred. No. 0.0017;  
Matches 230; Conservative 0; Mismatches 231; Indels 2; Gaps 1;

QY 444 aaaaactcaaaaataaactcctcgttttgaattgaactaagtgtttatttaca 503  
DB 60067 AAAAATACCAATCAGATTTTCTCTACGCTMAAATTAAGTACATCAACA 60008  
QY 504 ggtatgtaagatatatagatgaactgtttggaacacagatcgatcagctgatg 563  
DB 60007 GATACGGAGCTTCTTCAATCTCATCTTGAAGTTGTCACGTTAATCTCAATGGAAT 59948  
QY 564 ctgagactaatagatcatatatacaaaacgaagagcctcttgaatgtaatccta 623  
DB 59947 CAGAGTAAACAGATCAATTTAAAGAACGATCAAAAGCTTGCTGTTACCCAC 59888  
QY 624 gaagcatagtggaactcttggaatgagtcgactgtcttctgttgatgacatgata 683  
DB 59887 AATGATGCTTATATATCTTACTTGGAGCTTGATACATGCTGCGGTTACGCTGAGCCAC 59828  
QY 684 gatataggaagatctcgttaactcttgaatcagcagcagcttgaactattctac 743  
DB 59827 GCCTATATGAGAGCGCTGCTGCTCTCATATAGCTCGACCATGATGACATCAATCT 59768  
QY 744 ttaagaatgttgagaacatcttctgttctctgattcttgacacaaactcattt 803  
DB 59767 TGCCCTAAAGTTGATCATCAGAGAGCTATGATGATGAGAAAGCTTGAGGTTA 59708  
QY 804 tctcgtcacaagcagagccaaaagg--ttttatttatacctttatttgcataatt 861  
DB 59707 TTGATATCCAAATAGACCAACATGATGATATTTTATCACTTTTACATTTCTCTGT 59648  
QY 862 ttttcttataatcctttagagcttcaactttttttttt 904  
DB 59647 TTTACTTTTTTACCTTCCTGTTTTCAGCTACACACTGTT 59605

RESULT 13  
ATAP22 198354 bp DNA PLN 07-APR-1999  
LOCUS Arabidopsis thaliana DNA chromosome 4, ESSA I AP2 contig fragment  
DEFINITION No. 2.  
ACCESSION Z99708.1 GI:4006885  
VERSION Z99708.1  
KEYWORDS  
SOURCE  
ORGANISM  
thale cress.  
Arabidopsis thaliana

REFERENCE  
AUTHORS  
JOURNAL  
REFERENCE  
AUTHORS  
JOURNAL  
TITLE  
DIRECT SUBMISSION  
SUBMITTED (07-APR-1999) MIPS, at the Max-Planck-Institut fuer  
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG. E-mail:  
schoedle@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project  
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge

## COMMENT

Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK.  
E-mail: michael.bevan@jrc.ac.uk  
On Dec 12, 1998 this sequence version replaced gi:2464894.  
The annotation of this entry was produced with considerable  
contributions from Stephane Rombauts and Pierre Rouze, Department  
of Genetics, University of Ghent, Ledeganckstraat 35, 9000 Ghent,  
BE. E-mail: stromengens@rug.ac.be, p.rouze@rug.ac.be  
Information on performance of analysis and a more detailed  
annotation of this entry and other sequences of chromosomes 3, 4  
and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>  
This fragment has an overlap with ATAP21 at the 5' end.

## FEATURES

source  
1..198354  
/organism="Arabidopsis thaliana"  
/variety="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="4"  
1..10046  
/organism="Arabidopsis thaliana"  
/clone="cosmid C1C7A10.4E12"  
/db\_xref="taxon:3702"  
8165..10851  
/gene="C7A10.500"  
8165..8344  
/gene="C7A10.500"  
join(8165..8344,8417..8522,8638..8983,9116..9209,  
9502..9621,9720..9932,10027..10089,10166..10284,  
10372..10430,10523..10851)  
/gene="C7A10.500"  
/note="similar to hypothetical protein T517.13 from  
Arabidopsis thaliana chromosome II BAC T517 genomic  
sequence, PATY:G2642165  
contains Neutral zinc metalloproteinases, zinc-binding  
region signature (ILAHENMHM)  
contains EST gb:M65197, H76120, AA067496"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAB16816.1"  
/db\_xref="gi:4006885"  
/db\_xref="SPTREMBL:O23197"  
/translation="MEFLPLGCTIVFLVYAGKGLLPMGWLTFLTKSSHKESGQGN  
GRYEDRNLEGPRTSAGSDDEKEIECALALSLSEDEHYVPODGKRLLEYKSEIE  
EDDDDEDEDEEYKRAQLEAEAEERVAQAOIPEERKRAEAQLEFTEKLAKARLE  
EIEEMRRKAAQLEDELLAKALQESMANGSPRPYDGNLQYPLIPSSHICVGCQA  
EIGHGRFLSAWVARGILNVSAVMIHPNIPAGLIEYRAHPFMOKYCPSHEDGPRCC  
SCERMEPRKDYLLIDGRCLECLDSAIMDPEOCPLYLEIREYEGGLMKYQSN  
SYARGEISKRRAEGERHGHHLPETRGCLSEQYVTVLRRPRIGAGTKLIDMTT  
EPCRLIRCEVTAILILYGLPRLTGSILAHENMHMRLNGYPNLRPEVEEGICQVL  
AHMWLESETYAGSTLVDIASSSSAVVSSAKKGRSDFEKKLGEFFKHQIESDSSSA  
YGDGFRQGNQAVLKHGLRRLDHLRLGTFFPKW"  
8345..8416  
/gene="C7A10.500"  
/number=1  
8417..8522  
/gene="C7A10.500"  
/number=2  
8523..8637  
/gene="C7A10.500"  
/number=2  
8638..8983  
/gene="C7A10.500"  
/number=3  
8984..9115  
/gene="C7A10.500"  
/number=3  
9116..9209  
/gene="C7A10.500"  
/number=4  
9210..9501  
/gene="C7A10.500"  
/number=4  
9502..9621







exon	/gene="ATgq36190" /number=10 7653. .7736
intron	/gene="ATgq36190" /number=11 7737. .7853
exon	/gene="ATgq36190" /number=11 7854. .7894
intron	/gene="ATgq36190" /number=12 7895. .8795
exon	/gene="ATgq36190" /number=12 8796. .8921
intron	/gene="ATgq36190" /number=13 8922. .9011
exon	/gene="ATgq36190" /number=13 9012. .9185
intron	/gene="ATgq36190" /number=14 9186. .9264
exon	/gene="ATgq36190" /number=14 9265. .9450
intron	/gene="ATgq36190" /number=15 9451. .9551
exon	/gene="ATgq36190" /number=15 9552. .9668
intron	/gene="ATgq36190" /number=16 9669. .9767
exon	/gene="ATgq36190" /number=16 9768. .9821
intron	/gene="ATgq36190" /number=17 9822. .9900
exon	/gene="ATgq36190" /number=17 9901. .9972
intron	/gene="ATgq36190" /number=18 9973. .10582
exon	/gene="ATgq36190" /number=18 10583. .10657
intron	/gene="ATgq36190" /number=19 10658. .10726
exon	/gene="ATgq36190" /number=19 10727. .10865
intron	/gene="ATgq36190" /number=20 10866. .10979
exon	/gene="ATgq36190" /number=20 10980. .11095
intron	/gene="ATgq36190" /number=21 11360. .11432
exon	/note="Glu-tRNA (CTC) " 11536. .11982
intron	/gene="ATgq36200" complement(11536. .11982)
exon	/gene="ATgq36200" /number=1 complement(11536. .11982)

Query Match	Best Local Similarity	1.9%; Score 73.6; DB 13; Length 198750;
Matches 374; Conservative 0; Mismatches 384; Indels 30; Gaps 4;		
Query 491	tgattatttcaagtgatgtaagatatagatagatgaacttgcttgtagaaccacagatc	550
Ds 71942	tggttaacaaatpdtagatagcgaaagtggttcaaaacgaacatpdtataggacacacatcatp	71883
Qy 551	gtatcagctgtagctggaacttaatagatcaltacataacaaacgaagagctcttgtaa	610
Ds 71882	atatacmaacggatgacgagagtgaaatpaaotggtgcctccaaacacatgggacacatpdtg	71823
Qy 611	tgtagtatacctagaaagatagtggtgagcttttggaatgtagtgatctgtcttgct	670
Ds 71822	cctgcattaccccttaaatpcaatpaccgaactcttgagaaacctattctcagcatpcaat	71763
Qy 671	ggtgacatgataagatatgagaagatctcgcgttaactcttaagtcacagcagctct	730
Ds 71762	ggacctatctcaaaaaaggcttcacacgctcatpdtggcgggttcctcagatpctcaccctc	71703
Qy 731	agaactattctactaaagatgltgagagacatacttgcttgcttgatcttgagcaa	790
Ds 71702	aaagaccgagatcacgacgacatttgagagcctcggtgtgttcacactpdtggcgttgagct	71643
Qy 791	caaaactatcttctcctgcgtcaagaagagccaaagaggttttattatcttctat	850
Ds 71642	caactctcatggtgttcagatggttgagatagatcctcaaaaaaggtatgctcttagaaatpdt	71583
Qy 851	t-----ttgctaaatttttttgcttaagcttcttgtagagttctta	890
Ds 71582	tagtagcaaaactcacacgtacgtactagctacatgcctatpacaacttttttaattggta	71523
Qy 891	actttttttttttaaattgaacaglttaagtttaacttaactaalgcgaaagcatataagta	950
Ds 71522	gaatgcttgggtgtgtgtgtgtgagatgacggttgagattattagttaaagtggtgtagac	71463
Qy 951	atggaactctgtaagaagaagaacagagcaatlaaagaagaagtagtaacttctga	1010
Ds 71462	acatctcctg--TGAAGATATGAACAATTCCTCAAACTTGAAGATTCATCAAA	71406
Qy 1011	ggaagtgcctgcgtccctctaactctacacgaagctgttactcataagctctcagta	1070
Ds 71405	ggttgattttgtatcccaatcaaatpccctggcaccttaagactctacaaatpcccttaaaagta	71346
Qy 1071	catatttttttttgcgtgaagtcacaaactcattatagttttttaaatttatt	1130
Ds 71345	AA-----ATATTTCCTTATATACATGACGATATCATATCTGATTTCAATTTGTAGT	71292
Qy 1131	atggttataataaaat-cataaagtgtgtgtgtagtcaagagcaagaatltgaagtlc	1189
Ds 71291	gtgttaatactacagaggttttaatacttcaatcattgacgggaaagagaggttttaataaagtg	71232
Qy 1190	attgagagaaatgtaagagagaaatltgatatcaaggaagaagatcagaagaagaa	1249
Ds 71231	gtaaaaaaaggttgtagagagagacaaactggcagatgacaaacgactcttcggcacaatgac	71172
Qy 1250	gaagtga 1257	
Ds 71171	gtgctgga 71164	

```

XX 08-DEC-1997 (Rel. 53, Created)
DT 27-JUN-1998 (Rel. 56, Last updated, Version 2)
XX Plasmodium falciparum DNA *** SEQUENCING IN PROGRESS *** from contig 4-88
DE Plasmodium falciparum DNA *** SEQUENCING IN PROGRESS *** from contig 4-88
XX HTG.
XX Plasmodium falciparum (malaria parasite P. falciparum)
OS Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
XX
XX [1]
XX 1-21202
XX Lawson D.;
XX
XX Submitted (25-JUN-1998) to the EMBL/GenBank/DBJ databases.
XX P.falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome
XX Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK
XX
XX IMPORTANT: This sequence is unfinished and does not necessarily
XX represent the correct sequence. Work on the sequence is in progress
XX and the release of this data is based on the understanding that the
XX sequence may change as work continues. The sequence may be
XX contaminated with foreign sequence from E.coli, yeast, vector,
XX phage etc.
XX
XX For more information about this sequence or the Malaria Project,
XX see http://www.sanger.ac.uk/Projects/P\_falciparum.
XX
XX Key Location/Qualifiers
XX
XX source 1. 21202
XX /chromosome="4"
XX /db_xref="taxon:5833"
XX /organism="Plasmodium falciparum"
XX /clone="4-88"
XX
XX Sequence 21202 BP: 7427 A; 2248 C; 2061 G; 9466 T; 0 other;
SQ

```

```

Query Match 1.9%; Score 72.8; DB 41; Length 21202;
Best Local Similarity 45.5%; Pred. No. 0.021;
Matches 373; Conservative 0; Mismatches 442; Indels 5; Gaps 3;

```

```

OY 1523 ttagctattataagctatagatattctattatagatacgaatgaagaaagat 1582
DB 6380 TTGATTTTATTTATTTTATTTTATTTTGAAGTAAAGCAATATTTACAAAAA 6439
OY 1583 gtaatttaattgcaactaatttattgttattgatttacttgaagaaagacatc 1642
DB 6440 TTATATTTTATTTATTTATTTATTTATTTCTTTTAAATTTGTACCTTATAA 6499
OY 1643 ttgagatcgagagggccaaaggaactagagagcagaatlaaatggagattaca 1702
DB 6500 AAAAAACAAAAACAATATTTATTTGATTTTATTTGCTTTCTTATATCGTAAC 6559
OY 1703 agaaatgacttactcaatgtagtactatcatctt-cattattattctatgtt 1761
DB 6560 AAAAAAATAAACACACATTTTATGCTTTTATTTATTTATTTATTTAATCAT 6619
OY 1762 catagattatgagaaaccaaataatgatttttttttgggtgltgaaagttat 1821
DB 6620 TATTTTAAAGAAAAAACTTAAAGTTGATTTCTTTCTTTTAAATTAATTT 6679
OY 1822 aaatgaactcttcgtaggggaatgtagttagtlttgcacgcaagcaactcaaga 1881
DB 6680 TATATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 6739
OY 1882 ttttcgtagcaaaaggtaaaactttagcaaaatlttaataatgaatccggaatat 1941
DB 6740 GGGAAAAAATAAACAAATACATATTAATAAATAATTT--TTATTAATAATTC 6796
OY 1942 tgaactctattgtagtgaataataataataattacatttcttaacgttcgaa 2001

```

```

DB 6797 ACAATTTTATAAATTTTTCATTAATTTTCATTAATTTCCATTAATGCGATAAATA 6856
OY 2002 gatacagataccctagtgaggtaagtgtagcggtagctcagcgta-cattgat 2060
DB 6857 ATTAGACATTAAGGAATTTAAAAAATTTTAAAGAAATCTACAGTTATTAATTAAT 6916
OY 2061 aattctcgtagaccacctaactctttaaactcttgagagtgcaaggtlaataa 2120
DB 6917 TAAAAAATTTTAGAACAACTTTTAAAGAAAGACTGTTGTAATAAATAAATAA 6976
OY 2121 aagttctctcttactacacgaatagtgtagtlttttcaatcattgcatgaata 2180
DB 6977 AATTCAAAAAGAAAAAAGAGATTCATTTTAAATATTTATTTATTTATTTATTA 7036
OY 2181 gatacgtctacgtgatttactatctatagatacgaactaagaaactaactgcaa 2240
DB 7037 TATATATATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 7096
OY 2241 aaactttattattatattatcttcaagttagatcttcaacgctcattgcatgatca 2300
DB 7097 TTATTTAACATATTTATTTGAAATTAATTAATAAGACAAAAAATTAACATTA 7156
OY 2301 catgaagcatataataataataaataagagagagaaagaga 2340
DB 7157 CAAAAAATTAATAAATAACATATATTTATTTAGTTACATA 7196

```

Search completed: October 17, 2001, 19:48:49  
Job time: 18324 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2001, 19:18:28 : Search time 431.79 Seconds  
(without alignments)  
5655.317 Million cell updates/sec

Title: US-09-502-426-1\_COPY\_3000\_6888

Perfect score: 3889

Sequence: 1 atgaagctcattgtagtt.....cctagttcagccaaagctt 3889

Scoring table: IDENTITY\_MNC  
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.GeneSeq\_0601.\*  
1: /SIDS1/gcgdata/geneSeq/geneSeq/NA1980.DAT:\*  
2: /SIDS1/gcgdata/geneSeq/geneSeq/NA1981.DAT:\*  
3: /SIDS1/gcgdata/geneSeq/geneSeq/NA1982.DAT:\*  
4: /SIDS1/gcgdata/geneSeq/geneSeq/NA1983.DAT:\*  
5: /SIDS1/gcgdata/geneSeq/geneSeq/NA1984.DAT:\*  
6: /SIDS1/gcgdata/geneSeq/geneSeq/NA1985.DAT:\*  
7: /SIDS1/gcgdata/geneSeq/geneSeq/NA1986.DAT:\*  
8: /SIDS1/gcgdata/geneSeq/geneSeq/NA1987.DAT:\*  
9: /SIDS1/gcgdata/geneSeq/geneSeq/NA1988.DAT:\*  
10: /SIDS1/gcgdata/geneSeq/geneSeq/NA1989.DAT:\*  
11: /SIDS1/gcgdata/geneSeq/geneSeq/NA1990.DAT:\*  
12: /SIDS1/gcgdata/geneSeq/geneSeq/NA1991.DAT:\*  
13: /SIDS1/gcgdata/geneSeq/geneSeq/NA1992.DAT:\*  
14: /SIDS1/gcgdata/geneSeq/geneSeq/NA1993.DAT:\*  
15: /SIDS1/gcgdata/geneSeq/geneSeq/NA1994.DAT:\*  
16: /SIDS1/gcgdata/geneSeq/geneSeq/NA1995.DAT:\*  
17: /SIDS1/gcgdata/geneSeq/geneSeq/NA1996.DAT:\*  
18: /SIDS1/gcgdata/geneSeq/geneSeq/NA1997.DAT:\*  
19: /SIDS1/gcgdata/geneSeq/geneSeq/NA1998.DAT:\*  
20: /SIDS1/gcgdata/geneSeq/geneSeq/NA1999.DAT:\*  
21: /SIDS1/gcgdata/geneSeq/geneSeq/NA2000.DAT:\*  
22: /SIDS1/gcgdata/geneSeq/geneSeq/NA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1.	3889	100.0	6888	21	AA59599
C 2	200	5.1	936	22	AAF58252
C 3	200	5.1	936	22	AAF58254
C 4	200	5.1	936	22	AAF58257
C 5	200	5.1	936	22	AAF58259
C 6	200	5.1	936	22	AAF58262
C 7	200	5.1	936	22	AAF58255
C 8	198.8	5.1	936	22	AAF58252
C 9	198.8	5.1	936	22	AAF58254
C 10	198.8	5.1	936	22	AAF58257
C 11	198.8	5.1	936	22	AAF58259

#### ALIGNMENTS

12	198.8	5.1	936	22	AAF58262	Oligonucleotide D2
13	198.8	5.1	938	22	AAF58255	Oligonucleotide D1
14	92.8	2.4	4937	18	AAT85307	Arabidopsis thaliana
15	87.6	2.3	1608	18	AAT85306	Arabidopsis thaliana
16	87.6	2.3	1646	21	AAC48157	Arabidopsis thaliana
17	87.6	2.3	1649	21	AAC47986	Arabidopsis thaliana
18	71.4	1.8	244	22	AAF58238	Oligonucleotide D1
19	66.6	1.7	1398	21	AAC48714	Arabidopsis thaliana
20	66.2	1.7	244	22	AAF58238	Oligonucleotide D1
C 21	57.4	1.5	605	17	AAT31530	Human 3' apolipoprotein A-II
C 22	55.4	1.4	940	21	AAC37912	Arabidopsis thaliana
C 23	55.4	1.4	1673	21	AAC39858	Arabidopsis thaliana
C 24	54.2	1.4	665	21	AAZ31956	Human apolipoprotein A-II
C 25	52.8	1.4	49999	20	AAZ32902	Human LOBO homolog
C 26	52.6	1.4	2601	21	AAC48719	Arabidopsis thaliana
C 27	52.4	1.3	2300	20	AAZ00686	Human GPC3 DNA
C 28	52.2	1.3	1667	21	AAC35582	Arabidopsis thaliana
C 29	52.2	1.3	26811	20	AAZ20253	Borrelia burgdorferi
C 30	51.8	1.3	1697	21	AAC77781	Human cancer associated
C 31	51.6	1.3	1251	19	AAV34685	Arabidopsis thaliana
C 32	51.4	1.3	1493	21	AAZ43802	Human adult skin c
C 33	51.4	1.3	1733	21	AAZ37036	Human PRO1411 (UNQ
C 34	51.4	1.3	1733	22	AAZ54238	DNA encoding prote
C 35	51.4	1.3	1734	22	AAZ92083	Human PRO1411 cDNA
C 36	51.4	1.3	1734	22	AAZ87038	Nucleotide sequenc
C 37	51.2	1.3	2846	21	AAZ65034	Membrane-bound pro
C 38	51.2	1.3	2846	22	AAZ92076	Human PRO1344 cDNA
C 39	51.2	1.3	2846	22	AAZ44180	Human PRO1344 (UNQ
C 40	51	1.3	6644	20	AAZ33181	Base sequence of t
C 41	51	1.3	7372	20	AAZ33182	Base sequence of t
C 42	51	1.3	7797	20	AAZ33180	Cowpox virus bsr f
C 43	51	1.3	7996	20	AAZ33184	Base sequence of t
C 44	50.2	1.3	19124	18	AAT72682	Plasmidium var-7 g
C 45	50.2	1.3	19124	21	AAZ96287	Plasmidium var-7 p

#### RESULT 1

AA59599	standard; DNA; 6888 BP.
ID	AA59599
XX	AA59599;
AC	14-NOV-2000 (first entry)
XX	XX
DE	DNA encoding a cytochrome P450 enzyme designated DMF4.
KW	DMF4; cytochrome P450 enzyme; brassinosteroid; 22alpha-hydroxylation;
KW	plant phenotype; cell elongation; ss.
XX	XX
OS	Arabidopsis sp.
XX	XX
XX	XX
FT	key
FT	promoter
FT	Location/Qualifiers
FT	1..3203
FT	/*tag= a
FT	3060..3125
FT	/*tag= b
FT	3203..6110
FT	/*tag= c
FT	/product= "DMF4"
FT	/note= "contains introns"
FT	3203..3423
FT	/*tag= d
FT	3424..3503
FT	/*tag= e
FT	3504..3828
FT	/*tag= f
FT	3829..3913
FT	/*tag= g
FT	3914..4066
FT	/*tag= h
FT	exon

ET	Intron	4067..4164
ET	/tag- i	
ET	exon	4165..4479
ET	/tag- j	
ET	Intron	4480..4631
ET	/tag- k	
ET	exon	4632..4724
ET	/tag- l	
ET	Intron	4725..4815
ET	/tag- m	
ET	exon	4816..4894
ET	/tag- n	
ET	Intron	4895..5000
ET	/tag- o	
ET	exon	5001..5110
ET	/tag- p	
ET	Intron	5111..5864
ET	/tag- q	
ET	exon	5865..6110
ET	/tag- r	
ET	3'UTR	6011..6468
ET	/tag- s	
PN	WO20047715-A2.	
PD	17-AUG-2000.	
PE	11-FEB-2000; 2000WO-US03820.	
PR	11-FEB-1999; 99US-0119657.	
PR	11-FEB-1999; 99US-0119658.	
PA	(ARIZ-) ARIZONA BOARD OF REGENTS.	
PI	Azpirez R, Choe S, Feldmann KA;	
XX	WPI: 2000-549142/50.	
DR	P-PSDB; AAB07921.	
PT	New isolated dwf4 polynucleotide useful for altering the phenotype of plants, for diagnostic assays and in the production of antibodies -	
PS	Claim 3; Fig 10A-G; 113pp; English.	
XX	The present sequence encodes a DWF4 polypeptide. The polypeptide is a cytochrome P450 enzyme that mediates multiple steps in synthesis of brassinosteroids. Specifically, it mediates multiple	
CC	Z2alpha-hydroxylation steps in brassinosteroid biosynthesis. The DWF4 polynucleotide is used for altering the phenotype of a plant. DWF4 plants display a dramatic reduction in the length of different organs, and this size reduction is attributable to a defect in cell elongation.	
CC	The DWF4 polynucleotides and polypeptides can be used in diagnostic assays and to generate antibodies, which can be used to produce immunogenic compositions.	
SQ	Sequence 6888 BP: 2294 A; 1010 C; 1193 G; 2391 T; 0 other;	
XX		
Query Match	100.0%; Score 3889; DB 21; Length 6888;	
Best Local Similarity	100.0%; Pred. No. 0;	
Matches 3889; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
Db	1 atgaagccattgttgatgtaagttaagaataagaattttatatattttaatgaacat 60       	
OY	3000 atgaagcctatgggttgaagcttaataagaagatttttatattttaatgaacat 3059       	
Db	61 gataacaattatatttcctgaacttcctttaaaccctcttaacaagaacatccctt 120       	
OY	3060 gatacaaatatatttcttgactcttctttaaaccctcttcaacaagaagctccctt 3119       	
Db	121 ttcagtagaagtcgatcccaattcttaagaacaaggccattagaagaagtgaagtg 180       	
OY	3120 ttcaatagaagccgatcccatcttctaagaagaagccattagaagaagaagtgaagtg 3179       	

[illegible]

QY 1261 agagatgaacagagatgaatgaatgcatcgtgttagaacaacaagacacatga 1320  
|||||  
Db 4260 agagatgaacagagatgaatgaatgcatcgtgttagaacaacaagacacatga 4319  
QY 1321 tctttgggattgggttttgaaacatcgaattatcgacggagacaattctcgatcat 1380  
|||||  
Db 4320 tctttgggattgggttttgaaacatcgaattatcgacggagacaattctcgatcat 4379  
QY 1381 tcttagttgttatattgacggagatgaatcttctctgttagccattgtctgcacatct 1440  
|||||  
Db 4380 tcttagttgttatattgacggagatgaatcttctctgttagccattgtctgcacatct 4439  
QY 1441 ctctctgcaagcttgccctaaagccgttgaaagcttgaggtgaagatataaacaaga 1500  
|||||  
Db 4440 ctctctgcaagcttgccctaaagccgttgaaagcttgaggtgaagatataaacaaga 4499  
QY 1501 caagttaatctactacaacattgttgcgtatataatagaattatagaattatctatga 1560  
|||||  
Db 4500 caagttaatctactacaacattgttgcgtatataatagaattatagaattatctatga 4559  
QY 1561 gaataacagatgaanaaagtgtatattatattatgctactaatattatgattatgatt 1620  
|||||  
Db 4560 gaataacagatgaanaaagtgtatattatattatgctactaatattatgattatgatt 4619  
QY 1621 atactttgaaagaaagacatcttgatagatcgagagccaaagaaagaaactaagagatga 1680  
|||||  
Db 4620 atactttgaaagaaagacatcttgatagatcgagagccaaagaaagaaactaagagatga 4679  
QY 1681 gaattaaattgggattgattacaagaanaatggacttactcaatgtgtatgtactacat 1740  
|||||  
Db 4680 gaattaaattgggattgattacaagaanaatggacttactcaatgtgtatgtactacat 4739  
QY 1741 tctcaatttatattatgattatgattatgattatgattatgattatgattatgattatgatt 1800  
|||||  
Db 4740 tctcaatttatattatgattatgattatgattatgattatgattatgattatgattatgatt 4799  
QY 1801 ttgtgtgtgtgtgaaaggtatataaatagaacatctcgatgtggaattgtatgatttttt 1860  
|||||  
Db 4800 ttgtgtgtgtgtgaaaggtatataaatagaacatctcgatgtggaattgtatgatttttt 4859  
QY 1861 gcatcgcaaaagcactcaaaagtgtcgtgtacaaaggtaaacatttgcgtacaaaattttt 1920  
|||||  
Db 4860 gcatcgcaaaagcactcaaaagtgtcgtgtacaaaggtaaacatttgcgtacaaaattttt 4919  
QY 1921 aaataatgaataccggaattatgaaatcttattgtgataaataataataataatttaca 1980  
|||||  
Db 4920 aaataatgaataccggaattatgaaatcttattgtgataaataataataataatttaca 4979  
QY 1981 ttcttcaatgttggaaaaaagatacgataccctcagtcgggtgggaagtgttaccgggtga 2040  
|||||  
Db 4980 ttcttcaatgttggaaaaaagatacgataccctcagtcgggtgggaagtgttaccgggtga 5039  
QY 2041 tctcagccgtatcatttggataattctcgtgtacgacaaactaatcttcttaacttcttga 2100  
|||||  
Db 5040 tctcagccgtatcatttggataattctcgtgtacgacaaactaatcttcttaacttcttga 5099  
QY 2101 gatggcaacaggtataaanaagttctctcgtgttaactatcgaaaaatgagtataagttt 2160  
|||||  
Db 5100 gatggcaacaggtataaanaagttctctcgtgttaactatcgaaaaatgagtataagttt 5159  
QY 2161 ttctcatttgcattgataatgatacgtctcagtcgatttaccattatagataactataac 2220  
|||||  
Db 5160 ttctcatttgcattgataatgatacgtctcagtcgatttaccattatagataactataac 5219  
QY 2221 gagaactatcaatctgcaaaaaacttttattatattatattatcttccaagttagactttaa 2280  
|||||  
Db 5220 gagaactatcaatctgcaaaaaacttttattatattatattatcttccaagttagactttaa 5279  
QY 2281 acgtcaatgataatgatacgaatgaaagataataataaataaagaagagaagaga 2340  
|||||  
Db 5280 acgtcaatgataatgatacgaatgaaagataataataaataaagaagagaagaga 5339  
QY 2341 cgtgtgtgtgaagtgtacgtgaagaacaattatagtagagtagtattctttaaagac 2400  
|||||

Db 5340 cgtgtgtgtgaagtgtacgtgaagaacaattatagtagagtagtattctttaaagac 5399  
QY 2401 gtagagctgccttaataatcttctaatacgttgacggttgaattatattatagtcacgct 2460  
|||||  
Db 5400 gtagagctgccttaataatcttctaatacgttgacggttgaattatattatagtcacgct 5459  
QY 2461 tgatacaatttaagaatttgacggaagatgatacgcgcttgaaggtatctcaacgccc 2520  
|||||  
Db 5460 tgatacaatttaagaatttgacggaagatgatacgcgcttgaaggtatctcaacgccc 5519  
QY 2521 gatgactgtacgtgcttgaattctgcgaacgttgacggttgaattatattatagattata 2580  
|||||  
Db 5520 gatgactgtacgtgcttgaattctgcgaacgttgacggttgaattatattatagattata 5579  
QY 2581 actcattataataaacaactatcaaatcgcgaatttagagaataataactatataagatct 2640  
|||||  
Db 5580 actcattataataaacaactatcaaatcgcgaatttagagaataataactatataagatct 5639  
QY 2641 attagattattatagatgaataacttattgaaataagaataaataagtagatagatt 2700  
|||||  
Db 5640 attagattattatagatgaataacttattgaaataagaataaataagtagatagatt 5699  
QY 2701 atagtagtgagggaaggtgaagaagtggtgagagatgattatgacccacgtgtgtgt 2760  
|||||  
Db 5700 atagtagtgagggaaggtgaagaagtggtgagagatgattatgacccacgtgtgtgt 5759  
QY 2761 gccaaacagcagtgcttcttcttcttcttcttcttcccaactcttcttcttggggattat 2820  
|||||  
Db 5760 gccaaacagcagtgcttcttcttcttcttcttcttcccaactcttcttcttggggattat 5819  
QY 2821 ttgtattataaatacgtttgtcgttttttttttttttttttttttttttttttttttttttt 2880  
|||||  
Db 5820 ttgtattataaatacgtttgtcgttttttttttttttttttttttttttttttttttttttt 5879  
QY 2881 tcatcgtcaagagtgtagtttttctgcagcttgagggaacaaactacatgctgttttgagga 2940  
|||||  
Db 5880 tcatcgtcaagagtgtagtttttctgcagcttgagggaacaaactacatgctgttttgagga 5939  
QY 2941 gggccaaggtctatgtgtgttctcagagctcagccaagttagaagtcgaggtttatcat 3000  
|||||  
Db 5940 gggccaaggtctatgtgtgttctcagagctcagccaagttagaagtcgaggtttatcat 5999  
QY 3001 catctagttcttaaatcttaatttggaattagagaagaatgataccaacatttttcttct 3060  
|||||  
Db 6000 catctagttcttaaatcttaatttggaattagagaagaatgataccaacatttttcttct 6059  
QY 3061 ttgtgtgatttccctaaagcgtttgcctataggtttctcgtattctgtgaaaaaaaaaa 3120  
|||||  
Db 6060 ttgtgtgatttccctaaagcgtttgcctataggtttctcgtattctgtgaaaaaaaaaa 6119  
QY 3121 aagatgaagaatttttattctctctcttcttcttcttcttcttcttcttcttcttcttct 3180  
|||||  
Db 6120 aagatgaagaatttttattctctctcttcttcttcttcttcttcttcttcttcttcttct 6179  
QY 3181 cccaatgataataaanaatttggataaataataatattggataatctcgttttttaattctgg 3240  
|||||  
Db 6180 cccaatgataataaanaatttggataaataataatattggataatctcgttttttaattctgg 6239  
QY 3241 gtttgaanaaaggttctgacattcgaagtcgagagatgataatgattggagagtaggt 3300  
|||||  
Db 6240 gtttgaanaaaggttctgacattcgaagtcgagagatgataatgattggagagtaggt 6299  
QY 3301 tgaagcttggacatttgatgagatgtgtgtgattatagtgatgacacattaaacct 3360  
|||||  
Db 6300 tgaagcttggacatttgatgagatgtgtgtgattatagtgatgacacattaaacct 6359  
QY 3361 taaatgggcttctataagggcccaatttatataagataataaacaagtgacaaactttac 3420  
|||||  
Db 6360 taaatgggcttctataagggcccaatttatataagataataaacaagtgacaaactttac 6419  
QY 3421 ttctgttttgatcggaaacaaataatgtccaattgtccaattccaacaagaattatgataa 3480  
|||||

Db 6420 ttcgcttttgatccgaagcaataaacttgcataaaccacaagaattatgtaaa 6479  
Oy 3481 cactcgtgtgtctcagtggaatcatctggctgagactgaacatcagaacacaa 3540  
Db 6480 cactcgtgtgtctcagtggaatcatctggctgagactgaacatcagaacacaa 6539  
Oy 3541 acccttcaattatgataaacctcctatgacggttccaaacttatcttctatc 3600  
Db 6540 acccttcaattatgataaacctcctatgacggttccaaacttatcttctatc 6599  
Oy 3601 ggtgtacattgacacaaagatgagtgatcaaaagactcaaatgataatcctc 3660  
Db 6600 ggtgtacattgacacaaagatgagtgatcaaaagactcaaatgataatcctc 6659  
Oy 3661 aggataagaagtgatcaaaagatgagtgatcaaaagactcaaatgataatcctc 3720  
Db 6660 aggataagaagtgatcaaaagatgagtgatcaaaagactcaaatgataatcctc 6719  
Oy 3721 actgttcttgggtccattccttcaaatgagaaggttcttcttcttctatc 3780  
Db 6720 actgttcttgggtccattccttcaaatgagaaggttcttcttcttctatc 6779  
Oy 3781 atcactttagatcctccttctagacgagtaaatcctcgttaccagaagatgctc 3840  
Db 6780 atcactttagatcctccttctagacgagtaaatcctcgttaccagaagatgctc 6839  
Oy 3841 catctttgaagacaggtcttcccaagctcagtcagggccaaagctt 3889  
Db 6840 catctttgaagacaggtcttcccaagctcagtcagggccaaagctt 6888

## RESULT 2

AAFS8252/c  
ID AAF58252 standard; DNA; 936 BP.

AAFS8252:

24-APR-2001 (first entry)

Oligonucleotide D1835.

Electron-transfer group: ETM; mismatch; genotyping;

gene expression; ss.

Synthetic.

MO200107665-A2.

01-FEB-2001.

26-JUL-2000; 2000MO-US20476.

26-JUL-1999; 99US-0145695.

17-MAR-2000; 2000US-0190259.

(CLIN-) CLINICAL MICRO SENSORS INC.

Umek RM;

WPI: 2001-159728/16.

Nucleic acids containing electron-transfer group, useful as labels in hybridization assays, e.g. for genotyping, allowing repeat analyses on a single surface

Example 6; Page 127; 159pp; English.

The present invention relates to a composition comprising two nucleic acids each containing an electron-transfer group (ETM) having different redox potentials. The invention is used for electronic detection of nucleic acids, especially of substitutions (mismatches) and single-nucleotide polymorphisms, e.g. for genotyping, monitoring gene expression.

XX SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 5.1%; Score 200; DB 22; Length 936;  
Best Local Similarity 1.0%; Pred. No. 5,4e-28;  
Matches 8; Conservative 552; Mismatches 232; Indels 0; Gaps 0;

Oy 1479 gggtaagaataataaacgacgaagtaattaccacccaatgttgaagtataataag 1538  
Db 732 GGGCGCWW 733  
Oy 1539 ttattatgaattatcattagaaatacagatgaaaaglatatataattatgca 1598  
Db 732 WWW 673  
Oy 1599 ctaattatgttttcttattatcctttgaagaaagacatcgtgacgcgagggc 1658  
Db 672 WWW 613  
Oy 1659 caagaaggaactagagagtcagaatlaattggatgattacaagaatggacttac 1718  
Db 612 WWW 553  
Oy 1719 tcaatgtaatgtaatacattcattatattctatgltcatatgattatgata 1778  
Db 552 WWW 493  
Oy 1779 aaccaaaattatgttttttttgggtgtgtgtgaaggtataaatagaactctgat 1838  
Db 492 WWW 433  
Oy 1839 tgggaatgtagttagtttctgcatcgaagacactcaaatgttcgttacaagtgta 1898  
Db 432 WWW 373  
Oy 1899 aaactttagtaaaatlttaataatagaatccgaatatgaaatcttatgtgagtg 1958  
Db 372 WWW 313  
Oy 1959 aaaaatataataataatttaattcattccttaattgttgaaaaaagataatcctcgt 2018  
Db 312 WWW 253  
Oy 2019 gggtagaaagtgtaacggtgatacgcagcgtlacatttgataattcctgtagacca 2078  
Db 252 WWW 193  
Oy 2079 cctaactccttaactccttgagatgcaacaggtaaataaagttctcctgtaact 2138  
Db 192 WWW 133  
Oy 2139 atcgaaaattagtgatagtttttcatcatttgatgaatagatgaactcctcagtgatt 2198  
Db 132 WWW 73  
Oy 2199 tactatcatagatacactacgaaactaattatcgtgcaaaaactttattatattc 2258  
Db 72 WWW 13  
Oy 2259 atcttcaagtt 2270  
Db 12 WWW 1

## RESULT 3

AAFS8254/c  
ID AAF58254 standard; DNA; 936 BP.

AAFS8254:

24-APR-2001 (first entry)



DE Oligonucleotide D1875.  
XX Electron-transfer group; ETM; mismatch; genotyping;  
KM gene expression; ss.  
XX  
OS Synthetic.  
XX MO200107665-A2.  
PN  
PD 01-FEB-2001.  
XX  
PF 26-JUL-2000; 2000WO-US20476.  
XX  
PR 26-JUL-1999; 99US-0145695.  
PR 17-MAR-2000; 2000US-0190259.  
XX  
PA (CLIN-) CLINICAL MICRO SENSORS INC.  
XX  
PI Umek RM;  
XX  
DR WPI; 2001-159728/16.  
XX  
PT Nucleic acids containing electron-transfer group, useful as labels in  
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
PT a single surface  
XX  
PS Example 6; Page 127; 159pp; English.  
XX  
CC The present invention relates to a composition comprising two nucleic  
CC acids each containing an electron-transfer group (ETM) having  
CC different redox potentials. The invention is used for electronic  
CC detection of nucleic acids, especially of substitutions (mismatches)  
CC and single-nucleotide polymorphisms, e.g. for genotyping,  
CC monitoring gene expression.  
XX  
XX  
SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;  
  
Query Match 5.1%; Score 200; DB 22; Length 936;  
Best Local Similarity 1.0%; Pred.No.5.4e-28;  
Matches 8; Conservative 552; Mismatches 232; Indels 0; Gaps 0;  
  
QY 1479 gggtaagataataaacaagaagtaatactacccaatgttgcgtatataaag 1538  
DB 792 GGGGCGGWW 733  
QY 1539 ttattatagaattatctatagaataacgaatgaagaaatgtatattatga 1598  
DB 732 WWW 673  
QY 1599 ctaattatgtttatgtattatattacatttgaagaagaagcactgtgagtcgaagggc 1658  
DB 672 WWW 613  
QY 1659 caagaagaactagaagagtcagaatgaatggatgattacaagaagaatgacttac 1718  
DB 612 WWW 553  
QY 1719 tcaatgtatgtactacatcattcatattatctatgttcataatgatga 1778  
DB 552 WWW 493  
QY 1779 aaccaaatattgtttttttttgtgtgtgtggaagttataaactcttcgat 1838  
DB 492 WWW 433  
QY 1839 tgggaatagttagtttgcacgaagaacacgaatgttcggtacaaagta 1898  
DB 432 WWW 373  
QY 1899 aaacttcgtacaaatcttaataatgaatccggaatgtgaactctatgtgag 1958  
DB 372 WWW 313

QY 1959 aaatatataaataattacatttctaatgttgaaaaaagatacagatccctagt 2018  
DB 312 WWW 253  
QY 2019 ggggtgaagtgtagtcaggtacgcagcbacatttgatbaattctcgtatgacaa 2078  
DB 252 WWW 193  
QY 2079 cctaactcttaaccccttgagatgagcaacaggtaaataaagtctctcgttaact 2138  
DB 192 WWW 133  
QY 2139 atcgaataatagttagttttttcatctatgtcatgaatagatcagtcctcgtagt 2198  
DB 132 WWW 73  
QY 2199 taactatcatagatacactacgaagaactatbaatctgcaaaaactttatattat 2258  
DB 72 WWW 13  
QY 2259 atcttcaagtt 2270  
DB 12 WWWWWWWWWWWWW 1  
  
RESULT 4  
AAF58257/C  
ID AAF58257 standard; DNA; 936 BP.  
XX  
XX AAF58257;  
AC  
XX 24-APR-2001 (first entry)  
DT  
XX  
DE Oligonucleotide D1954.  
XX  
XX Electron-transfer group; ETM; mismatch; genotyping;  
KM gene expression; ss.  
XX  
OS Synthetic.  
XX  
PN WO200107665-A2.  
PD 01-FEB-2001.  
XX  
PF 26-JUL-2000; 2000WO-US20476.  
XX  
PR 26-JUL-1999; 99US-0145695.  
PR 17-MAR-2000; 2000US-0190259.  
XX  
PA (CLIN-) CLINICAL MICRO SENSORS INC.  
XX  
PI Umek RM;  
XX  
DR WPI; 2001-159728/16.  
XX  
PT Nucleic acids containing electron-transfer group, useful as labels in  
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
PT a single surface  
XX  
PS Example 6; Page 127; 159pp; English.  
XX  
CC The present invention relates to a composition comprising two nucleic  
CC acids each containing an electron-transfer group (ETM) having  
CC different redox potentials. The invention is used for electronic  
CC detection of nucleic acids, especially of substitutions (mismatches)  
CC and single-nucleotide polymorphisms, e.g. for genotyping,  
CC monitoring gene expression.  
XX  
SQ Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;  
  
Query Match 5.1%; Score 200; DB 22; Length 936;

[illegible][illegible]

QY 1479 gggtaaagataattataacagcacaaagttaattactaccaatgttcgtattataag 1538

PN  
XX  
XX  
WOZ001U/865-AZ.

2  
1  
1  
2  
2  
2  
1



```

QY 1602 attatatttatttatttatttatttgaagaagacatcttgagatcgagggccaa 1661
DB 121 ..... 180
QY 1662 gaaggaactagagagatcgaaatattgtagatgattacaagaatgacttactaa 1721
DB 181 ..... 240
QY 1722 atgtgtatgtactatcatcttatttatttatttatttatttatttatttattt 1781
DB 241 ..... 300
QY 1782 caaatatttatttatttatttatttatttatttatttatttatttatttattt 1841
DB 301 ..... 360
QY 1842 gaagtgtatgtatttatttatttatttatttatttatttatttatttatttattt 1901
DB 361 ..... 420
QY 1902 cttacgtacaaatttttaataatgaaatccggaatttgaatcttattgattgaa 1961
DB 421 ..... 480
QY 1962 aatatataataatttatttatttatttatttatttatttatttatttatttattt 2021
DB 481 ..... 540
QY 2022 tggaaagtgtacccgtgatactcagccgtacatttggataattcgttataccaact 2081
DB 541 ..... 600
QY 2082 aatcccttaacccttgtagatgacacaggttaataaatttctcctgtaatactac 2141
DB 601 ..... 660
QY 2142 gaaaattagttatagtttttcatcatctgcatgatatgacgtccatgatttac 2201
DB 661 ..... 720
QY 2202 ctatctatagatactatacgaagaactaatactctgcaaaaactttattattatc 2261
DB 721 ..... 780
QY 2262 ttccaa 2267
DB 781 ..... 786

```

## RESULT 9

```

AAF58254
ID AAF58254 standard; DNA; 936 BP.
XX
AC AAF58254;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1875.
XX
KN Electron-transfer group; ETM; mismatch; genotyping;
XX gene expression; ss.
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000MO-US20476.
XX
PR 26-JUL-1999; 980US-0145695.
PR 17-MAR-2000; 2000US-0190259.

```

```

XX (CLIN-) CLINICAL MICRO SENSORS INC.
XX
XX Umek RM;
XX
XX WPI; 2001-159728/16.
XX
XX Nucleic acids containing electron-transfer group, useful as labels in
XX hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX a single surface.
XX
XX Example 6; Page 127; 159pp; English.
XX
XX The present invention relates to a composition comprising two nucleic
XX acids each containing an electron-transfer group (ETM) having
XX different redox potentials. The invention is used for electronic
XX detection of nucleic acids, especially of substitutions (mismatches)
XX and single-nucleotide polymorphisms, e.g. for genotyping,
XX monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

```

Query Match 5.1%; Score 198.8; DB 22; Length 936;  
 Best Local Similarity 0.6%; Pred. No. 8.9e-28;  
 Matches 5; Conservative 552; Mismatches 229; Indels 0; Gaps 0;

```

QY 1482 taagataattataacagcaagtaataactaccatgttgaatgattatataagta 1541
DB 1 ..... 60
QY 1542 ttatagaattatttatttgaataatacgaatgaataatgattatatttatttacta 1601
DB 61 ..... 120
QY 1602 atttattgttatttatttatttatttatttatttatttatttatttatttattt 1661
DB 121 ..... 180
QY 1662 gaaggaactagagagatcgaaatattgtagatgattacaagaatgacttactaa 1721
DB 181 ..... 240
QY 1722 atgtgtatgtactatcatcttatttatttatttatttatttatttatttatttattt 1781
DB 241 ..... 300
QY 1782 caaatatttatttatttatttatttatttatttatttatttatttatttattt 1841
DB 301 ..... 360
QY 1842 gaagtgtatgtatttatttatttatttatttatttatttatttatttatttattt 1901
DB 361 ..... 420
QY 1902 cttacgtacaaatttttaataatgaaatccggaatttgaatcttattgattgaa 1961
DB 421 ..... 480
QY 1962 aatatataataatttatttatttatttatttatttatttatttatttatttattt 2021
DB 481 ..... 540
QY 2022 tggaaagtgtacccgtgatactcagccgtacatttggataattcgttataccaact 2081
DB 541 ..... 600
QY 2082 aatcccttaacccttgtagatgacacaggttaataaatttctcctgtaatactac 2141
DB 601 ..... 660
QY 2142 gaaaattagttatagtttttcatcatctgcatgatatgacgtccatgatttac 2201
DB 661 ..... 720

```

Db 661 www..... 720  
Qy 2202 ctatctatagatacactagagaactataatctcggcaaacctttttattatc 2261  
Db 721 www..... 780  
Qy 2262 ttccaa 2267  
Db 781 www... 786

## RESULT 10

AAF58257  
ID AAF58257 standard; DNA; 936 BP.

XX AAF58257;

XX 24-APR-2001 (first entry)

DE Oligonucleotide D1954.

KM Electron-transfer group; ETM; mismatch; genotyping;

KW gene expression; ss.

OS Synthetic.

PN WO200107665-A2.

PD 01-FEB-2001.

PF 26-JUL-2000; 2000WO-US20476.

PR 26-JUL-1999; 99US-0145695.

PR 17-MAR-2000; 2000US-0190259.

XX (CLIN-) CLINICAL MICRO SENSORS INC.

PI Umek RM;

DR WPI: 2001-159728/16.

PT Nucleic acids containing electron-transfer group, useful as labels in  
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
PT a single surface

PS Example 6; Page 127; 159pp; English.

XX The present invention relates to a composition comprising two nucleic  
CC acids each containing an electron-transfer group (ETM) having  
CC different redox potentials. The invention is used for electronic  
CC detection of nucleic acids, especially of substitutions (mismatches)  
CC and single-nucleotide polymorphisms, e.g. for genotyping,  
CC monitoring gene expression.

XX Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match 5.1%; Score 198.8; DB 22; Length 936;

Best Local Similarity 0.6%; Pred. No. 8.9e-28;

Matches 5; Conservative 552; Mismatches 229; Indels 0; Gaps 0;

Qy 1482 taagataattataacagacaagttaattactccaattgtacgtattataagta 1541

Db 1 www..... 60

Qy 1542 ttatagaattatctatagaatacagatgaaaaagatgtatattatgtcacta 1601

Db 61 www..... 120

Qy 1602 atttaattatgtattatactttgaaggaagacatcttgacgcgagggccaa 1661

Db 121 www..... 180

Qy 1662 gaaggactagagagatcgaattaaattggaatgattacaaagaatgacttactca 1721

Db 181 www..... 240

Qy 1722 atgtatgtactacatctcatattatctcatgttcatatgattatgataaac 1781

Db 241 www..... 300

Qy 1782 caaatattgattttttttgtgtgtgtgaaagtataaataaaccttgattgg 1841

Db 301 www..... 360

Qy 1842 geaatgtatgtatttctgcacgaagcactcaagaatgtctgcgtacaaggtaaa 1901

Db 361 www..... 420

Qy 1902 cttaacgtacaaattttaaataaagaalccggaatatgaattatgtatgaa 1961

Db 421 www..... 480

Qy 1962 aatatataaattatcatcttcttaagtgtgaaaaaagatacagatccctagtgg 2021

Db 481 www..... 540

Qy 2022 tggaaagtgtaccggtgactcagcgtacatttgataattcgtatgaccaact 2081

Db 541 www..... 600

Qy 2082 aatctcttaaccttgagatgcaacaggtlaaataaagttctcgttaactatc 2141

Db 601 www..... 660

Qy 2142 gaaattgtatagttttttcatctatgtcatgaatagatagacgtcctagtttac 2201

Db 661 www..... 720

Qy 2202 ctatctatagatacactagagaactataatctgcaaaaactttattatc 2261

Db 721 www..... 780

Qy 2262 ttccaa 2267

Db 781 www... 786

## RESULT 11

AAF58259  
ID AAF58259 standard; DNA; 936 BP.

XX AAF58259;

XX 24-APR-2001 (first entry)

DE Oligonucleotide D2004.

XX Electron-transfer group; ETM; mismatch; genotyping;

KW gene expression; ss.

OS Synthetic.

PN WO200107665-A2.

PD 01-FEB-2001.

PF 26-JUL-2000; 2000WO-US20476.

PR 26-JUL-1999; 99US-0145695.

PR 17-MAR-2000; 2000US-0190259.

XX (CLIN-) CLINICAL MICRO SENSORS INC.

PI Umek RM;



```

Db 241 ..... 300
Qy 1782 caaatatgtttttttgtgtgtgaagtttaaatgaacttcgttg 1841
Db 301 ..... 360
Qy 1842 gaatgttagttagtttgcgaagcactcaagatgttcgttaagaagtaaa 1901
Db 361 ..... 420
Qy 1902 cttacgtacaaaattttaaatgaatccgaataatgaatctatgtgaa 1961
Db 421 ..... 480
Qy 1962 aatataaataatttacctctctaagtgtgaaaaaagtaagatccctgtgg 2021
Db 481 ..... 540
Qy 2022 tggaaagtgtacggtgacgcgcgtacatttgataatctcgtatgacaact 2081
Db 541 ..... 600
Qy 2082 aatccttaactccttgagatggaacaggtlaaataaagttctcgttaactc 2141
Db 601 ..... 660
Qy 2142 gaaatagtgatagttttcatcactatgcatgataagtcctcaagtgttac 2201
Db 661 ..... 720
Qy 2202 ctactatagatactacagaaactatcctgcgaataactttatattatc 2261
Db 721 ..... 780
Qy 2262 ttcaaa 2267
Db 781 ..... 786

```

RESULT 13  
AAFS8255  
ID AAF58255 standard; DNA; 938 BP.  
XX  
AC AAF58255;  
XX  
DT 24-APR-2001 (first entry)  
XX  
DE Oligonucleotide D1876.  
XX  
KM Electron-transfer group; ETM; mismatch; genotyping;  
XX  
KW gene expression; ss.  
XX  
OS Synthetic.  
XX  
PN WO200107665-A2.  
XX  
PD 01-FEB-2001.  
XX  
PF 26-JUL-2000; 2000WO-US20476.  
XX  
PR 26-JUL-1999; 99US-0145695.  
XX  
PR 17-MAR-2000; 2000US-0190259.  
XX  
PA (CLIN-) CLINICAL MICRO SENSORS INC.  
XX  
PI Umek RM;  
XX  
DR WPI; 2001-159728/16.  
XX  
PT Nucleic acids containing electron-transfer group, useful as labels in  
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
PT a single surface

```

XX Example 6; Page 127; 159pp; English.
PS
XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;

Query Match 5.1%; Score 198.8; DB 22; Length 938;
Best Local Similarity 0.6%; Pred. No. 8.9e-28;
Matches 5; Conservative 552; Mismatches 229; Indels 0; Gaps 0;

Qy 1482 taagataattataacagcaagtaattactaccaaattgtacgtatataagta 1541
Db 1 ..... 60
Qy 1542 ttatgaattatctcattgaataacgataagaaaaagtatcatattatcacta 1601
Db 61 ..... 120
Qy 1602 atttatgttatgtattatattactttgaaggaagacacttgatcgagggcaa 1661
Db 121 ..... 180
Qy 1662 gaagaaactagagagtcagaataatggatgattcaagaataatgacttactca 1721
Db 181 ..... 240
Qy 1722 atgtgtagtactatcattccattattatcctatgttcaatgatattagatgaac 1781
Db 241 ..... 300
Qy 1782 caaatatgtttttttgtgtgtgaagtttaaatgaacttcgttg 1841
Db 301 ..... 360
Qy 1842 gaatgttagttagtttgcgcgaagcactcaagatgttcgttaagaagtaaa 1901
Db 361 ..... 420
Qy 1902 cttacgtacaaaattttaaatgaatccgaataatgaatctatgtgaa 1961
Db 421 ..... 480
Qy 1962 aatataaataatttacctctctaagtgtgaaaaaagtaagatccctgtgg 2021
Db 481 ..... 540
Qy 2022 tggaaagtgtacggtgacgcgcgtacatttgataatctcgttaagcaact 2081
Db 541 ..... 600
Qy 2082 aatccttaactccttgagatggaacaggtlaaataaagttctcgttaactc 2141
Db 601 ..... 660
Qy 2142 gaaatagtgatagttttcatcactatgcatgataagtcctcaagtgttac 2201
Db 661 ..... 720
Qy 2202 ctactatagatactacagaaactatcctgcgaataactttatattatc 2261
Db 721 ..... 780
Qy 2262 ttcaaa 2267
Db 781 ..... 786

```



RESULT 14  
AAT85307  
ID AAT85307 standard; DNA; 4937 BP.  
XX  
AC AAT85307;  
XX  
DT 14-APR-1998 (first entry)  
XX  
DE Arabidopsis thaliana cytochrome P450-type hydroxylase DNA.  
XX  
KW Cytochrome P450-type hydroxylase; identification; brassinosteroid;  
KW brassinosteroid inhibitor; modified plant; recombinant production;  
KW teasterone; ds.  
XX  
OS Arabidopsis thaliana.  
XX  
FH Key Location/Qualifiers  
FT CDS 968..4773  
FT /tag= a  
FT /product= cytochrome\_P450-type\_hydroxylase  
FT intron 1484..1679  
FT /tag= b  
FT /number= 1  
FT intron 1830..1916  
FT /tag= c  
FT /number= 2  
FT intron 2166..3902  
FT /tag= d  
FT /number= 3  
FT intron 3990..4083  
FT /tag= e  
FT /number= 4  
FT intron 4163..4247  
FT /tag= f  
FT /number= 5  
FT intron 4355..4445  
FT /tag= g  
FT /number= 6  
FT intron 4577..4673  
FT /tag= h  
FT /number= 7  
XX  
PN W09735986-A1.  
XX  
PD 02-OCT-1997.  
XX  
PE 27-MAR-1997; 97WO-EP01586.  
XX  
PR 27-MAR-1996; 96US-0622166.  
XX  
PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.  
XX  
PI Altmann T, Koncz C, Mathur J, Szekeres MA;  
XX  
DR WPI: 1997-489649/45.  
XX  
DR P-PSDB: AAM27153.  
XX  
PT New isolated plant cytochrome P450-type hydroxylase gene - used to  
PT identify substances acting as brassino-steroid(s) or brassinosteroid  
PT inhibitors for the production of modified plants  
XX  
PS Claim 1; Pages 48-53; 77pp; English.  
XX  
XX The present sequence encodes Arabidopsis thaliana cytochrome  
CC P450-type hydroxylase. The hydroxylase can be used to identify  
CC brassinosteroids or brassinosteroid inhibitors, useful to produce  
CC plants with modified physiological and/or phenotypic  
CC characteristics. The modified plants may show, e.g. stimulation of  
CC growth, increased cell elongation, increased wood production, in  
CC accelerated seed germination at low temperatures, an increase in  
CC dry weight, repressed anthocyanin production during growth in light  
CC and/or inhibited de-etiolation which is induced, e.g. by cytokinin,

CC in the dark or an increase in stress tolerance. The hydroxylase or  
CC its coding sequence can also be used for the recombinant production  
CC of compounds, e.g. teasterone.  
XX  
SQ Sequence 4937 BP; 1526 A; 864 C; 870 G; 1677 T; 0 other;  
  
Query Match 2.4%; Score 92.8; DB 18; Length 4937;  
Best Local Similarity 53.1%; Pred. No. 2.9e-08;  
Matches 225; Conservative 0; Mismatches 192; Indels 7; Gaps 1;  
  
QY 502 caggtatgtgaagatataagatcgaactgttttggaagacccaagatcgtatcagctga 561  
DB 1162 cccgtaacgagcgttcttcacatgacgcacatcttttggtaacgagatcttctcagcga 1221  
QY 562 tgcgtgaacttaagatcattacaaacgaaggaagcgtcttgaatgattatcc 621  
DB 1222 cccggaacgaacggttcttcttcacagaacgaaggaagcgttcttgaatgattatcc 1281  
QY 622 tagaagataggttggtggttcttgggaagatggtcgtctgttctgttgatcga 681  
DB 1282 tgcctccattgttaacctttggggaaacacctcgtcttcttaagaaaggtcttgca 1341  
QY 682 tagagatatgaagaaatctcgtcctaacttcttaagtcacgcgctctagaactatct 741  
DB 1342 taaacgtatgacactctctccacatgacgttcttgaatcttcaatcaatgaacacatct 1401  
QY 742 acttaagaatgttgagagacatacttctgttcttcttgatcttgcgaacaaactctat 801  
DB 1402 catgcttgatgacgcgttgatgcgttgatcttgaatcttcttctgttcttctc 1454  
QY 802 ttctctgtctcaagacgagcccaaaagttttatttttactttattttgtctaact 861  
DB 1455 gttctctccctcattggaagaagcaaaagtaacaaataatcttctgttctcaaaac 1514  
QY 862 ttttcttattgaactcttagagttcttaacttttttttttaattgaacagtttaag 921  
DB 1515 attatatattattttatttcgctctcactcactaagttttttttaataataaaataaat 1574  
QY 922 ttaa 925  
DB 1575 aaaa 1578  
  
RESULT 15  
AAT85306  
ID AAT85306 standard; cDNA to mRNA; 1608 BP.  
XX  
AC AAT85306;  
XX  
DT 14-APR-1998 (first entry)  
XX  
DE Arabidopsis thaliana cytochrome P450-type hydroxylase cDNA.  
XX  
KW Cytochrome P450-type hydroxylase; identification; brassinosteroid;  
KW brassinosteroid inhibitor; modified plant; recombinant production;  
KW teasterone; ds.  
XX  
OS Arabidopsis thaliana.  
XX  
FH Key Location/Qualifiers  
FT CDS 48..1467  
FT /tag= a  
FT /product= cytochrome\_P450-type\_hydroxylase  
XX  
PN W09735986-A1.  
XX  
PD 02-OCT-1997.  
XX  
PE 27-MAR-1997; 97WO-EP01586.  
XX  
PR 27-MAR-1996; 96US-0622166.  
XX

PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN

PI Altmann T, Koncz C, Mathur J, Szekeres MA, ...

DR WPI; 1997-489649/45.  
DR P-PSDB; AAW27153.

DR P-PSDB; AAW27153.

PT New isolated plant cytochrome P450-type hydroxylase gene - used to  
PT identify substances acting as brassino-steroid(s) or brassinosteroid  
PT inhibitors for the production of modified plants

PT inhibitors for the production of modified plants

PT inhibitors for the production of modified plants

PS Claim 1; Pages 44-46; 77pp; English.

The present sequence encodes Arabidopsis thaliana cytochrome P450-type hydroxylase. The hydroxylase can be used to identify brassinosteroids or brassinosteroid inhibitors, useful to produce plants with modified physiological and/or phenotypic characteristics. The modified plants may show, e.g., stimulation of growth, increased cell elongation, increased wood production, accelerated seed germination at low temperatures, an increase in dry weight, repressed anthocyanin production during growth in light and/or inhibited de-etiolation which is induced, e.g., by cytokinin, in the dark or an increase in stress tolerance. The hydroxylase or its coding sequence can also be used for the recombinant production of compounds, e.g., testosterone.

CC of compounds, e.g. testosterone.

CC of compounds, e.g. testosterone.

CC of compounds, e.g. testosterone.

CC of compounds, e.g. testosterone.

CC of compounds, e.g. testosterone.

CC of compounds, e.g. testosterone.

CC of compounds, e.g. testosterone.

CC of compounds, e.g. testosterone.

CC of compounds, e.g. testosterone.

CC of compounds, e.g. testosterone.

CC of compounds, e.g. testosterone.

SQ Sequence 1608 BP; 382 A; 374 C; 374 G; 478 T; 0 other;

Query Match	Score	DB	Length
2.38;	87.6;	18;	1608;

Best Local Similarity 56.6%; Pred. No. 2.1e-07;  
Matches 162; Conservative 0; Mismatches 124

Matches 162; Conservative 0; Mismatches 124; Indels 0; Gaps 0.

QY 502 caggtatgtaagatatagatcgaacttglttgagaaccaacgatacgtatcagctga 561

Db 242 ccggtacggttcggtttcattgacgcattcttttggtgaaccgacgatttctcagctga 301

QY 562 tgctgacttaatagattcatattacaacgaagagctcttgaatgtagtattcc 621

Db 302 ccggaacgaaccgtttgttctctcagaacgaagggaagcttttgagtgttcttatcc 361

QY 622 tagaagtataggtgggattcttgggaaatggtcgatgtctgttctgtgtgtgacatgca 681

Db 362 tgcctccattgtaaccttctgggaaacactctctgcttcttatgaaagttcttgc 421

QY 682 tagagatatgagaagtatctcgcttaacttcttaagtcacgcacgctcttagaactattct 741

Db 422 taaacgtatgcactctctcaccatgagctttgtctaattcttcaatcatctaagaccatct 481

QY 742 acttaagatgttgagagacatacttgttcttctgattcttg 787

Db 482 catgctgatatgacggttagtcggtttaatcttgaattcttg 527

Search completed: October 17, 2001, 19:19:13  
Job time: 15953 sec

Job time: 15953 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2001, 19:13:08 ; Search time 156 Seconds

(without alignments)  
4719.429 Million cell updates/sec

Title: US-09-502-426-1\_COPY\_3000\_6888

Perfect score: 3889

Sequence: 1 atgaagcatcgttgtagtt.....cctagtcagcgcaagctt 3889

Scoring table: IDENTITY\_NUC

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCIOS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfillseq1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92.8	2.4	4937	2	US-08-622-166A-3
2	87.6	2.3	1608	2	US-08-622-166A-1
3	62.6	1.6	7218	1	US-08-232-463-14
4	54.8	1.4	837	4	US-08-998-416-288
5	54.2	1.4	665	2	US-08-883-795A-36
6	54	1.4	636	4	US-08-998-416-1137
7	51.6	1.3	1251	4	US-08-972-927-7
8	50.2	1.3	19124	2	US-08-487-826B-13
9	48.8	1.3	615	2	US-08-998-416-186
10	47.6	1.2	665	2	US-08-883-795A-36
11	47	1.2	1461	5	US-08-722-126A-4
12	47	1.2	1461	5	PCR-US95-04258-4
13	46.8	1.2	3401	1	US-08-218-943-4
14	46.4	1.2	1117	4	US-09-247-373B-33
15	46.2	1.2	518	1	US-08-485-284A-2
16	45.6	1.2	5852	1	US-07-867-106-2
17	45.4	1.2	467	2	US-08-841-349-18
18	45.4	1.2	6671	1	US-08-280-443-1
19	45.4	1.2	6671	1	US-08-457-451-1
20	45.4	1.2	6671	1	US-08-555-678-1
21	45.4	1.2	1474	5	PCR-US95-02275-1
22	45	1.2	1641	1	US-08-821-994-64
23	45	1.2	1641	1	US-08-300-903A-8
24	44.8	1.2	1066	1	US-08-157-101A-4
25	44.8	1.2	1134	3	US-09-248-335-29
26	44.8	1.2	1882	4	US-09-370-253-1
27	44.8	1.2	2082	2	US-08-785-310A-2

28	44.8	1.2	3701	4	US-08-845-258-10	Sequence 10, Appl
29	44.8	1.2	3701	4	US-08-990-571-10	Sequence 10, Appl
30	44.6	1.1	6243	2	US-09-056-075-1	Sequence 1, Appl
31	44.2	1.1	658	4	US-08-998-416-595	Sequence 595, App
32	44.2	1.1	1736	3	US-09-182-816-22	Sequence 22, Appl
33	44.2	1.1	1736	3	US-09-182-816-24	Sequence 24, Appl
34	44.2	1.1	1736	3	US-09-471-528-22	Sequence 22, Appl
35	44.2	1.1	1736	3	US-09-471-528-24	Sequence 24, Appl
36	44	1.1	8220	2	US-08-568-459A-11	Sequence 11, Appl
37	44	1.1	8220	2	US-08-487-826B-11	Sequence 11, Appl
38	43.8	1.1	998	4	US-09-122-400B-5	Sequence 5, Appl
39	43.8	1.1	1850	3	US-08-617-860B-32	Sequence 32, Appl
40	43.8	1.1	3095	6	5231168-1	Patent No. 5231168
41	43.8	1.1	4098	2	US-08-605-106-4	Sequence 4, Appl
42	43.6	1.1	1098	3	US-09-248-335-35	Sequence 35, Appl
43	43.6	1.1	1534	1	US-08-300-903A-6	Sequence 6, Appl
44	43.6	1.1	2550	6	5258287-23	Patent No. 5258287
45	43.6	1.1	2671	6	5168051-9	Patent No. 5168051

#### ALIGNMENTS

RESULT 1  
US-08-622-166A-3  
; Sequence 3, Application us/08622166A  
; Patent No. 5952545  
; GENERAL INFORMATION:  
; APPLICANT: KONCZ, CSABA  
; APPLICANT: MATYUR, JAIDEEP  
; APPLICANT: SZEKERES, MIKLOS  
; APPLICANT: ALTMANN, THOMAS  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING CYTOCHROME  
; NUMBER OF SEQUENCES: 4  
; P450-TYPE PROTEINS INVOLVED IN THE BRASSINOSTEROID SYNTHESIS  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP  
; STREET: P.O. Box 747  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/622,166A  
; FILING DATE: 27-MAR-1996  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SVENSSON, LEONARD R.  
; REGISTRATION NUMBER: 30,330  
; REFERENCE/DOCKET NUMBER: 0147-0153P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 205-8000  
; TELEFAX: (703) 205-8050  
; TELEX: 248345  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4937 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Arabidopsis thaliana  
; STRAIN: cv. Columbia  
; IMMEDIATE SOURCE:  
; LIBRARY: lambda gt10

```

CLONE: C204
FEATURE:
NAME/KEY: CDS
LOCATION: join(968..1483, 1680..1829, 1917..2165, 3903
LOCATION: ..3989, 4084..4162, 4248..4354, 4446..4576, 4674
LOCATION: ..4773)
US-08-622-166A-3

```

```

Query Match          2.4% Score 92.8; DB 2; Length 4937;
Best Local Similarity 53.1%; Pred. No. 1.5e-10;
Matches 225; Conservative 0; Mismatches 192; Indels 7; Gaps 1;

```

```

OY 502 caggtatgtaagatagatatagatcgaactgttttgagaagcaacagatcagtcga 561
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1162 CCGGTACGGTGGTTCATGACGATCTTTTGGTGAACCGAGATTTCTCAGCGTA 1221
OY 562 tgcctgactaataatcattacaaacgaaggaagcctcttgatgtatcc 621
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1222 CCGGAAACGACCGTTCCTTCAGAACGAGGAGGATTTTGAAGTCTTATCC 1281
OY 622 tagaagtagtgagtgatctcttgagaagtcgagtcgtctgtgtgtgagatga 681
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1282 TCCCTTCATTTTAACCTTTTGGGAAACACTCTCTGCTTTATGAAGGTTCTTGA 1341
OY 682 tagagatatagaagtagtcgttaactcttaagtcacgacgctctagaactatct 741
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1342 TAAAGGTATGACCTCTCACCATGAGCTTTCGTAATCTTCAATCATTAAGACCATCT 1401
OY 742 acttaagaatgttgagagacataactgttctgtctgtatctcttgcaacaactctat 801
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1402 CATGCTGTATATGACCGGTTAGTCCGGTTTAATCTGATTCATTCG-----TCTTCTC 1454
OY 802 ttctctctcaagacgagcgcaaaagttttattttatctttatcttttgcataact 861
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1454 GGTCTCTCTCATGGAAGACCAAAAGGTAAACCAAAAATTTCTGCTTATCAAAAAC 1514
OY 862 ttctgttataatccttaagatcttaacttttttttttttttttttttttttttttt 921
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1515 ATTATATATATTATTTTATGCGCTTCACATTATGTTTATTAATAAATAAATAAT 1574
OY 922 ttaa 925
    ||
DB 1575 AAAA 1578

```

```

RESULT 2
US-08-622-166A-1
Sequence 1, Application US/08622166A
Patent No. 5952545
GENERAL INFORMATION:
APPLICANT: KONCZ, CSABA
APPLICANT: MATUR, JASDEEP
APPLICANT: SZEKRES, MIKLOS
APPLICANT: ALTMANN, THOMAS
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING CYTOCHROME
NUMBER OF INVENTION: P450-TYPE PROTEINS INVOLVED IN THE BRASSINOSTEROID SYNTHESIS I
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/622,166A
FILING DATE: 27-MAR-1996

```

```

CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: SVENSSON, LEONARD R.
REGISTRATION NUMBER: 30,330
REFERENCE/DOCKET NUMBER: 0147-0153P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1608 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: A. thaliana
IMMEDIATE SOURCE:
LIBRARY: lambda gt10
CLONE: C204
FEATURE:
NAME/KEY: CDS
LOCATION: 48..1466
US-08-622-166A-1

```

```

Query Match          2.3% Score 87.6; DB 2; Length 1608;
Best Local Similarity 56.6%; Pred. No. 1.3e-09;
Matches 162; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

```

```

OY 502 caggtatgtaagatagatatagatcgaactgttttgagaagcaacagatcagtcga 561
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 242 CCGGTACGGTGGTTCATGACGATCTTTTGGTGAACCGAGATTTCTCAGCGTA 301
OY 562 tgcctgactaataatcattacaaacgaaggaagcctcttgatgtatcc 621
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 302 CCGGAAACGACCGGTTTGTTCCTTCAGACGAGGAGGAGGATTTTGAAGTCTTATCC 361
OY 622 tagaagtagtgagtgatctcttgagaagtcgagtcgtctgtctgtgtgtgagatga 681
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 362 TCCCTTCATTTTAACCTTTTGGGAAACACTCTCTGCTTTATGAAGGTTCTTGA 421
OY 682 tagagatatagaagtagtcgttaactcttaagtcacgacgctctagaactatct 741
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 422 TAAAGGTATGACCTCTCACCATGAGCTTTCGTAATCTTCAATCATTAAGACCATCT 481
OY 742 acttaagaatgttgagagacataactgttctgtctgtatctcttg 787
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 482 CATGCTGTATATGACCGGTTAGTCCGGTTTAATCTGATTCATTCG 527

```

```

RESULT 3
US-08-232-463-14/C
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent-In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZypt-flis
US-08-232-463-14

```

Query Match	1.68	Score 62.6	DB 1	Length 7218
Best Local Similarity	9.78	Pred. NO. 0.00029		
Matches 26	Conservative 151	Mismatches 90	Indels 0	Gaps 0

[illegible]

RESULT 4  
 US-08-998-416-288  
 : Sequence 288, Application US/08998416  
 : Patent No. 6239264  
 :  
 : GENERAL INFORMATION:  
 : APPLICANT: Philippsen, Peter  
 : APPLICANT: Pohlmann, Rainer  
 : APPLICANT: Steinert, Sabine  
 : APPLICANT: Mohr, Christine  
 : APPLICANT: Wendland, Jurgen  
 : APPLICANT: Knechtle, Philipp  
 : APPLICANT: Reibischung, Corinne  
 : TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPII  
 : TITLE OF INVENTION: AND USES THEREOF  
 : NUMBER OF SEQUENCES: 1152  
 :  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: NO. 6239264artis Corporation

```

1 STREET: 3054 Cornwallis Road
2 CITY: Research Triangle Park
3 STATE: No. 6239264th Carolina
4 COUNTRY: USA
5 ZIP: 27709
6
7 COMPUTER READABLE FORM:
8 MEDIUM TYPE: Floppy disk
9 COMPUTER: IBM PC compatible
10 OPERATING SYSTEM: PC-DOS/MS-DOS
11 SOFTWARE: PatentIn Release #1.0, Version #1.30
12 CURRENT APPLICATION DATA:
13 APPLICATION NUMBER: US/08/998,416
14 FILING DATE: 24-DEC-1997
15 CLASSIFICATION: 435
16
17 PRIOR APPLICATION DATA:
18 APPLICATION NUMBER: CH 0016/97
19 FILING DATE: 31-DEC-1996
20 ATTORNEY/AGENT INFORMATION:
21 NAME: Meigs, J. Timothy
22 REGISTRATION NUMBER: 36,241
23 REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
24 TELECOMMUNICATION INFORMATION:
25 TELEPHONE: 919-541-8587
26 TELEFAX: 919-541-8689
27
28 INFORMATION FOR SEQ ID NO: 288:
29 SEQUENCE CHARACTERISTICS:
30 LENGTH: 837 base pairs
31 TYPE: nucleic acid
32 STRANDEDNESS: single
33 TOPOLOGY: linear
34
35 MOLECULE TYPE: DNA (genomic)
36 ORIGINAL SOURCE:
37 ORGANISM: PAG1241RP
38
39 US-08-998-416-288

```

MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:  
; ORGANISM: PAG1241RP  
US-08-998-416-288

Query Match	1.4%	Score 54.8	DB 4	Length 837
Best Local Similarity	44.4%	Pred. No. 0.0066		
Matches 221	Conservative 0	Mismatches 277	Indels 0	Gaps
QY 1521	tggtacgtaataataagattatataagaattatcttcattcagatatataacatgataaaaagt	1580		
DB 75	TTTAACTATTATATCATATTTTAAATTAATTAATTAATTTGATTTTAATCTATTATTA	134		
QY 1581	atgatataatcgtacataatttctatctgattatgattataactcttgaaagaaagca	1640		
DB 135	TAAATTAATTAATTAATTTACTTAATTCATCATATTAATTAATTAATTAATTAATTA	194		
QY 1641	tcctgagatcgcgaagcccaagaagaactcaggagagctcagatataatttggatgata	1700		
DB 195	TATTTAATATGCAATACTATTTACTCTATGTCGCAAAATTTTAAATTTAGTTATTAATTA	254		
QY 1701	caagaaatgagacttcaactcaatcgtatgatactcaatctcattatcttaactatgt	1760		
DB 255	TTAGCTATTATATTCTTTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	314		
QY 1761	tcatatgattatgatagaagaaacaaatatactgattcttcttctggcgtgtaaggtta	1820		
DB 315	TTTATTATATGTTTATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTA	374		
QY 1821	taaatgaacactctcggttgggaagatgtaggttttctgatacgaaagactcaag	1880		
DB 375	TATTTGTAATTAATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTA	434		
QY 1881	atgttcggtataaagaagaaacttcaactgatacaaaattttaataataagaatccgata	1940		
DB 435	TTATTTTATCTTTTATAAGAATTAATTAATTAATTAATTAATTAATTAATTAATTA	494		
QY 1941	ttgaaactctatgatagaagaaatataaataatatacttcaacttctgaatgttgaaaaa	2000		
DB 495	TTAATTTTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	554		
QY 2001	gqatacagataccctaact	2018		



US-08-972-927-7/c  
Sequence 7, Application US/08972927  
Patent No. 6166290  
GENERAL INFORMATION:  
APPLICANT: Rea, Philip A  
APPLICANT: Lu, Yu-ping  
APPLICANT: Li, Ze-sheng  
TITLE OF INVENTION: GLUTATHIONE-S-CONJUGATE TRANSPORT IN  
PLANTS  
NUMBER OF INVENTIONS: 20  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.  
STREET: One Commerce Square, 2005 Market Street, 22nd  
STREET: Floor  
CITY: Philadelphia  
STATE: Pennsylvania  
COUNTRY: US  
ZIP: 19103-7086  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/972,927  
FILING DATE: 18-NOV-1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/031,040  
FILING DATE: 18-NOV-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/061,328  
FILING DATE: 08-OCT-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Doyle Leary Ph.D., Kathryn  
REGISTRATION NUMBER: 36,317

OY	746	aagaatgttgagaaacatacttggtttgttcctgcatccttgcaacaacctatttc	805
Dd	994	AAATGGGTAATACTGTATCTTTTTCATATTAGCAACTAATACATTAAGTTTGTA	935
OY	806	tctgcataagacagaggccaaaaggtttatatttatcttattcttgtaattttc	865
Dd	934	TTCCTTACAATAATTTATATTATTAATTCATTATCTGCTATTTCTCATAATTCAT	875
OY	866	tgtttaagaactttagagtttctaactttttttttttaaattgaacagtttaagttaa	925
Dd	874	CTTTTATTTAANAAGCTCATCCCAATTTTTTACTATATTTTGAAGAAAATAAATAATATA	815
OY	926	tctaattcgagaacatalaatgtagttagtccttggaagaagaacacagacaaattaa	985
Dd	814	GGTATTTTAAAAATGCATATCTAGAAGAGAGAGTTGGATATATATATCGTCGCCAAGTTAA	755
OY	986	gaaagagatgatcaacttccaatgaagaagatgtctctgcctcctaaalctaacgagac	1045
Dd	754	CMAACATTTTTTCTTTTGGAAGAAAGGATATTGATTTATCGTCAGGAAAATTTGGAATC	695
OY	1046	tgctatcataaagccctccaaagtlacatatatttttctgcgtgaagaagtcgaacct	1105
Dd	694	AAGTATTTCTACATATGGTTAAATCTCTTCTTGACACCTTAGTATTTTTTAACTTTAT	635
OY	1106	cattatagtttcaattattattcttagtltgaataaatactcaaaatgtgttgtag	1165
Dd	634	TAAACTATCCAATGMAAATCATGTCCTTTAAAAAACACATGATTTTAGTAGATGATCATTA	575
OY	1166	tcacga	1171
Dd	574	TCACGA	569

RESULT 8  
 US-08-487-826B-13/C  
 : Sequence 13 Application US/08487826B  
 : Patent No. 5993827  
 :  
 : GENERAL INFORMATION:  
 :  
 : APPLICANT: Sim, Kim L.  
 : APPLICANT: Chlunis, Chetan  
 : APPLICANT: Miller, Louis H.  
 : APPLICANT: Peterson, David S.  
 : APPLICANT: Su, Xin-zhaun  
 : APPLICANT: Wellens, Thomas E.  
 : TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
 : TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
 : NUMBER OF SEQUENCES: 45  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Knobbe Martens Olson & Bear  
 : STREET: 620 Newport Center Drive 16th Floor  
 : CITY: Newport Beach  
 : STATE: California  
 : COUNTRY: US  
 : ZIP: 92660

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,826B  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelsen, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH121.001CP1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19124 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
US-08-487-826B-13

Query Match 1.3% Score 50.2; DB 2; Length 19124;

Best Local Similarity 43.0%; Pred. No. 0.14; Mismatches 538; Indels 9; Gaps 3;  
Matches 412; Conservative 0;

QY 1506 taattactacaatgttgcgtatataatgaattatagaattatctatagaata 1565  
DB 6945 TAATAAATTAAATTTCTAACAAAAAAATATATACAGAAATATATATTATGTA 6886  
QY 1566 taagctgaaaaaagatgataatctaatgtcacaatttatgttcatcact 1625  
DB 6885 TTCCCTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAAG 6826  
QY 1626 ttgtagaagaagcgtcttgagatcgagggcgaagaaggaaggaagtcagcaatt 1685  
DB 6825 TAAATTTATTAATGAGAAAAAATACGAAATATACAAATATTAATAATATATAG 6766  
QY 1686 aaattggagatgatacaagaatgtacttcaatgctgact--atcattc 1742  
DB 6765 CAACGTGTTTATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 6706  
QY 1743 tcaattatctatgttcataatgataatgaatgaatgaatgaatgaatgaatgaat 1802  
DB 6705 TTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 6646  
QY 1803 tgggtgtgtagaaggttaataatgaactctgacttgtagtgtagtgtagtgtagtgc 1862  
DB 6645 TTCTCTATCATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 6586  
QY 1863 atcgcaaaagcactcaaaagatgttcgtaaaaggaaggaaggaaggaaggaagga 1922  
DB 6585 TACATATTTCTTTGTATTTGTATTTATTTAAAGTACTATTTATTTATTTATTT 6526  
QY 1923 ataataaagccgaatctgaatcttattgagatgaataatataatataatataat 1982  
DB 6525 GTTGTATATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 6466  
QY 1983 tcttaattgtgaaaaaagatagatataccctagtggttgtagaaggttttaaccgtgac 2042  
DB 6465 AATAATATACAGAAAAATATATTTATTTATTTATTTATTTATTTATTTATTTA 6406  
QY 2043 tcaagcgtacatttgtagaattctcgtaagcaactaactaactaactaactaacta 2102  
DB 6405 TTATCATTTGTTTGTGATTTTATTTATTTATTTATTTATTTATTTATTTATTT 6346  
QY 2103 tggcacaagcgaataaaaaagttctcgttaactatcgaataatagtgatagtttt 2162

DB 6345 TATGTAT 6286  
QY 2163 tcaattatgcatgataatgataatgataatgataatgataatgataatgataatgata 2219  
DB 6285 TGAGTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 6226  
QY 2220 cgagacttaataatcgtgcaaaacttttattatataatcgtgcaatgataatgata 2279  
DB 6225 AT 6169  
QY 2280 caagtcagatcattgataatgataatgataatgataatgataatgataatgataatgata 2339  
DB 6168 AAAAT 6109  
QY 2340 acgtgtgtgtagaagtgtagaaggaaggaaggaaggaaggaaggaaggaaggaagga 2399  
DB 6108 AT 6049  
QY 2400 cgtagagcgtgcaataatcttataatcgtgacgctgattatagtcagcgcg 2458  
DB 6048 TATTCCAATATGATATTTGTTTATTTATTTATTTATTTATTTATTTATTTATTT 5990

RESULT 9  
US-08-998-416-186  
Sequence 186, Application US/08998416  
Patent No. 6239264  
GENERAL INFORMATION:  
APPLICANT: Philippsen, Peter  
APPLICANT: Pohlmann, Rainer  
APPLICANT: Steiner, Sabine  
APPLICANT: Mohr, Christine  
APPLICANT: Wendland, Jürgen  
APPLICANT: Knechtle, Philipp  
APPLICANT: Redlschunig, Corinne  
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPYII  
NUMBER OF SEQUENCES: 1152  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6239264artis Corporation  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: No. 6239264th Carolina  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/998,416  
FILING DATE: 24-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: CH 0016/97  
FILING DATE: 31-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 186:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 615 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:







PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/799,479  
FILING DATE: 14-NOV-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 71007/118/USGO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3401 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-218-943-4

Query Match 1.2%; Score 46.8; DB 1; Length 3401;  
Best Local Similarity 50.4%; Pred. No. 0.44;  
Matches 114; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

Oy 845 ttttatttgcctaattttttgttgaatcctttagagttcctaattttttttt 904  
Db 561 ttttttcaagcgtggagctttgatttgacagtttgcattgttattgtttgttgc 640  
Oy 905 aattgaacagttacgttcaatcctaatgacgaacataatgataatgataatgcctgaga 964  
Db 641 agttaaaggaaatctttaaataattttgaacaggttaagaanaatccttgatcaacatg 700  
Oy 965 agaaagaacagagcaatcaagaagaagatgtaacttcaatgaagaaggtgtctgcgc 1024  
Db 701 caatgatcagagaaatgaagtcacagatattttattgttattgttggaggaactctgc 760  
Oy 1025 tcccttaactcaccaggaactgctatcataaagctcttcagta 1070  
Db 761 accatattattgttccacacttgatgatttgcattccatgcagga 806

RESULT 14  
US-09-247-373B-33/C  
Sequence 33, Application US/09247373B  
Patent No. 6168854  
GENERAL INFORMATION:  
APPLICANT: MCGONIGLE, BRIAN  
APPLICANT: O'KEEFE, DANIEL  
TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE ENZYMES  
FILE REFERENCE: CL-1108-A  
CURRENT APPLICATION NUMBER: US/09/247,373B  
CURRENT FILING DATE: 1999-02-10  
PRIOR APPLICATION NUMBER: 08/924,747  
PRIOR FILING DATE: 1997-09-05  
NUMBER OF SEQ ID NOS: 56  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 33  
LENGTH: 1117  
TYPE: DNA  
ORGANISM: SOYBEAN  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1101)  
OTHER INFORMATION: M-A OR C  
NAME/KEY: unsure  
LOCATION: (1104)  
OTHER INFORMATION: M-A OR C  
NAME/KEY: unsure  
LOCATION: (1116)  
OTHER INFORMATION: M-G OR A OR T OR C  
US-09-247-373B-33

Query Match 1.2%; Score 46.4; DB 4; Length 1117;  
Best Local Similarity 65.4%; Pred. No. 0.39;  
Matches 68; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Oy 831 ttttatttlaacttcttatttgcctaatttttgccttaacatcctaagtttcta 890  
Db 1074 ttt 1015  
Oy 891 actttttttttttaaattgaacagttacgttcaatcctaagtcgc 934  
Db 1014 tttttttttttttttttttttttaaagaattgaattgattttgtcc 971

RESULT 15  
US-08-485-284A-2  
Sequence 2, Application US/08485284A  
Patent No. 5750372  
GENERAL INFORMATION:  
APPLICANT: SAKAI, YASUYOSHI  
APPLICANT: TANI, YOSHIKI  
APPLICANT: SHIBANO, YUJI  
APPLICANT: KONDO, HIROTO  
APPLICANT: HATANAKA, HARUYO  
TITLE OF INVENTION: NOVEL VECTOR HAVING PROMOTER THAT IS  
INDUCIBLE BY METHANOL AND/OR GLYCEROL  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESS: CUSHMAN DARBY & CUSHMAN, L.L.P.  
STREET: 1100 New York Avenue, N.W.  
CITY: Washington  
STATE: D. C.  
COUNTRY: U.S.A.  
ZIP: 20005-3918  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,284A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: A35  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 43361/1992  
FILING DATE: 28-FEB-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/025,416  
FILING DATE: 01-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: WHITE JR., PAUL E.  
REGISTRATION NUMBER: 32,011  
REFERENCE/DOCKET NUMBER: 217755/FPS382090S  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 518 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: genomic DNA  
US-08-485-284A-2

Query Match 1.2%; Score 46.2; DB 1; Length 518;  
Best Local Similarity 54.4%; Pred. No. 0.35;  
Matches 93; Conservative 0; Mismatches 78; Indels 0; Gaps 0;  
Oy 3048 attgttctccttattgtgatttctcaacggttgcctataaggttctcgtattct 3107  
||||| ||||| | ||||| ||| ||| ||||| ||| |||

```

Db      54  ATTTTTCATCAATTTATTTATTTTTCACCTTTTAAATCTTGATATGAT 113
Oy      3108  gtaaaaaaaagatgaagatatttattccctcttttttgataatttaa 3167
Db      114  ATGATATGATATGATTTAGTCTTGCTGTGTTTTTTTTTTTTCACAACTTTCTT 173
Oy      3168  atcatttttttgcccaatgatalataaaaaatttgataataatataattatt 3218
Db      174  TTAATGACTTTATACCAAAATTTCAAAAATTTCCAAAAAACCAAT 224

```

Search completed: October 17, 2001, 19:14:07  
 Job time: 15722 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2001, 16:25:00 ; Search time 5433.22 Seconds  
(without alignments)  
6766.181 Million cell updates/sec

Title: US-09-502-426-1\_COPY\_3000\_6888  
Perfect score: 3889  
Sequence: 1 atgagctcattggttagt.....cctagtcagccaagctt 3889

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 segs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est33:\*  
26: gb\_est34:\*  
27: gb\_est35:\*  
28: gb\_est36:\*  
29: gb\_est37:\*  
30: gb\_est38:\*  
31: gb\_est39:\*  
32: gb\_est40:\*  
33: em\_estba:\*  
34: em\_estfun:\*  
35: em\_esthum1:\*  
36: em\_esthum2:\*  
37: em\_esthum3:\*  
38: em\_esthum4:\*  
39: em\_esthum5:\*  
40: em\_esthum6:\*  
41: em\_esthum7:\*  
42: em\_esthum8:\*  
43: em\_esthum9:\*

44: em\_esthum10:\*  
45: em\_esthum11:\*  
46: em\_esthum12:\*  
47: em\_esthum13:\*  
48: em\_esthum14:\*  
49: em\_esthum15:\*  
50: em\_esthum16:\*  
51: em\_esthum17:\*  
52: em\_esthum18:\*  
53: em\_esthum19:\*  
54: em\_esthum20:\*  
55: em\_esthum21:\*  
56: em\_esthum22:\*  
57: em\_esthum23:\*  
58: em\_esthum24:\*  
59: em\_esthum25:\*  
60: em\_esthum26:\*  
61: em\_esthum27:\*  
62: em\_esthum28:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estin5:\*  
68: em\_estom1:\*  
69: em\_estom2:\*  
70: em\_estov1:\*  
71: em\_estov2:\*  
72: em\_estp11:\*  
73: em\_estp12:\*  
74: em\_estp13:\*  
75: em\_estp14:\*  
76: em\_estp15:\*  
77: em\_estp16:\*  
78: em\_estp17:\*  
79: em\_estp18:\*  
80: em\_estp19:\*  
81: em\_estp110:\*  
82: em\_estro1:\*  
83: em\_estro2:\*  
84: em\_estro3:\*  
85: em\_estro4:\*  
86: em\_estro5:\*  
87: em\_estro6:\*  
88: em\_estro7:\*  
89: em\_estro8:\*  
90: em\_estro9:\*  
91: em\_estro10:\*  
92: em\_estro11:\*  
93: em\_estro12:\*  
94: em\_estro13:\*  
95: em\_estro14:\*  
96: em\_estro15:\*  
97: em\_estro16:\*  
98: em\_estro17:\*  
99: em\_estro18:\*  
100: em\_estro19:\*  
101: em\_estro20:\*  
102: gb\_est25:\*  
103: gb\_est26:\*  
104: gb\_est27:\*  
105: gb\_est28:\*  
106: gb\_est29:\*  
107: gb\_est30:\*  
108: gb\_est31:\*  
109: gb\_est32:\*  
110: gb\_est41:\*  
111: gb\_est42:\*  
112: gb\_est43:\*  
113: gb\_est44:\*  
114: gb\_est45:\*  
115: gb\_est46:\*  
116: gb\_est47:\*

```
117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*
190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



BE315722	BE315722	642 bp	mRNA	EST	14-JUL-2000
LOCUS	BE315722				
DEFINITION	NF025D11LF1F1090. Developing leaf Medicago truncatula cDNA clone				
ACCESSION	NF025D11LF 5 , mRNA sequence.				
VERSION	BE315722				
KEYWORDS	BE315722.1 GI:9189499				
SOURCE	EST.				
ORGANISM	barrel medic.				
	Medicago truncatula				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;				
	Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.				
REFERENCE	1 (bases 1 to 642)				
AUTHORS	Torres-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,				
	Flores,H.R., Inman,J.T., Weller,J.W. and May,G.D.				
TITLE	Expressed Sequence Tags from the Samuel Roberts Noble Foundation				
JOURNAL	Medicago truncatula leaf library				
COMMENT	Unpublished (2000)				
	Contact: May GD				

Plant Biology Division  
The Samuel Roberts Noble Foundation  
2510 Sam Noble Parkway, Ardmore, OK 73402, USA  
Tel: 580 221 7391  
Fax: 580 221 7380  
Email: [gdmay@noble.org](mailto:gdmay@noble.org)  
Medicago Genome Initiative accession: MGI:S:16068  
Insert Length: 642 Std Error: 0.00  
Plate: 025 row: D column: 11  
Seq primer: TCACACAGCAACAGCTATGAC.

FEATURES	Location/Qualifiers
source	1. .642

```

/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="NF02SD11LF"
/clone_id="Developing leaf"
/lssue_type="leaf"
/dev_stage="Pooled developmental"
/node_vector="lambda Zap: Contains a mixture of very
young, developing, mature and senescing leaves."
BASE COUNT      210 a      110 c      109 g      213 t
ORIGIN

```

Query Match	4.88;	Score 186.8;	DB 166;	Length 642;
Best Local Similarity	76.28;	Pred. No. 2.4e-25;		
Matches 230;	Conservative 0;	Mismatches 72;	Indels 0;	Gaps 0

[illegible][illegible]

Public Soybean EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [est@watson.wustl.edu](mailto:est@watson.wustl.edu)  
This clone is available through: Genome Systems, Inc., 4633 World  
Parkway Circle St. Louis, Missouri 63134 For further information  
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)  
427-3324 or contact: [clones@genomesystems.com](mailto:clones@genomesystems.com) or  
[info.genomesystems.com](http://info.genomesystems.com) web site: [www.genomesystems.com](http://www.genomesystems.com)  
High quality sequence stop: 420.

FEATURES	Location/Qualifiers
source	1. .667

```

/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1066-1781"
/clone_id="Gm-c1066"
/tissue_type="leaf and shoot tip, salt stressed, 2 week
old seedling"
/lab_host="DH10B"
/notes="Vector: pluscript II SK+, site_1: EcoRI, site_2:
XhoI; The cDNA library was constructed from mRNA isolated
from unexpanded leaves and the shoot tips of 2 week old
seedlings from the cultivar Williams. The 2 week old
seedlings were salt stressed in a solution of 500mM NaCl
for 3 days prior to harvesting. Complementary DNA was
synthesized from mRNA using a primer consisting of a
poly(dT) sequence with a XhoI restriction site. EcoRI
adapters were ligated to the blunt-ended cDNA fragments
followed by XhoI digestion. The cDNA fragments were
directionally cloned into the EcoRI-XhoI restriction site
of the pluscript vector. The ligated cDNA fragments were
transformed into DH10B host cells (GibcoBRL). This library
was constructed in the laboratory of Dr. Randy
Shoemaker."

```

BASE COUNT	194	a	155	c	141	g	177	t
ORIGIN								
Query Match	4.6%; Score 178.6; DB 147; Length 667;							
Best Local Similarity	71.4%; Pred. No. 8.3e-24;							
Matches 235; Conservative	0; Indels 0; Gaps 0;							

[illegible]



Db 278 GCAGAGCATCAAGAGGTTCTACAAAGAGGAATGTCAGTGCAGCTACTCT 337  
 Qy 623 aagaatgatggggatctcttggaagaatgctgatgctgtcttgcttgatcatgcat 682  
 Db 338 ACAAGCATCGTGGAATCTAGGAATAATGTCATGTCCTTACTGATGACATGAT 397  
 Qy 683 aagaatgatgaagatctcgcgttaactcttaagtcacgacgctcttaagatctca 742  
 Db 398 AGAGACATCGGGTTTATCCTCACTCAACTTCTTAGCCACGCCAGCTCGAACAACCTTC 457  
 Qy 743 cctaaagatgttgagagacatcttgcttgcttgatcttgctgcaacaaactctat 802  
 Db 458 TTGAAGAGGTGAGAGAGATCCCTCTGCTGATCTTGAACCTTGAGCCCAAGTTGCATA 517  
 Qy 803 tctctgctcaagagagagccaaaagt 831  
 Db 518 TTCTCAGCCCAAGATGAAGTAAAGATT 546

RESULT 4  
 LOCUS BF050501 533 bp mRNA EST 16-OCT-2000  
 DEFINITION EST435659 tomato developing/immature green fruit Lycopersicon  
 esculentum cDNA clone CLEM18G22 5' sequence, mRNA sequence.  
 ACCESSION BF050501  
 VERSION BF050501.1 GI:10804397  
 KEYWORDS EST.  
 SOURCE tomato.  
 ORGANISM Lycopersicon esculentum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots;  
 Asteridae; euasterids I; Solanales; Solanales; Solanaceae; Solanum;  
 Lycopersicon.  
 REFERENCE 1 (bases 1 to 533)  
 AUTHORS Alcalá, J., Vrebalov, J., White, R., van der Hoeven, R.S., Holt, I.E.,  
 Liang, F., Hansen, T.S., Craven, M.B., Bowman, C.L., Rønning, C.M.,  
 Niemann, W., Fraser, C.M., Martin, G.B., Giovannoni, J.J. and Tanksley,  
 S.D.  
 TITLE Generation of ESTs from tomato fruit tissue, immature green  
 Unpublished (2000)  
 JOURNAL Contact: David Frisch  
 COMMENT Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Tel.: 864 656 4366  
 Fax: 864 656 4293  
 Email: dfrisch@CLEMSON.EDU.  
 FEATURES  
 source 1..533  
 Location/Qualifiers  
 /organism="Lycopersicon esculentum"  
 /cultivar="TA496"  
 /db\_xref="taxon:4081"  
 /clone\_lib="CLEM18G22"  
 /clone\_lib="tomato developing/immature green fruit"  
 /tissue\_type="fruit"  
 /dev\_stage="immature green (5-35 days post-anthesis)"  
 /lab\_host="SOLR"  
 /note="Vector: pBluescriptSKmCudaapt; Site\_1: EcoRI;  
 Site\_2: XhoI; Fruit were tagged at 5 dpa (0.5 cm) and  
 harvested at 7 day intervals through 35 dpa. Equal masses  
 of tissue from each stage were combined (including seeds  
 and locules) prior to mRNA isolation."  
 BASE COUNT 172 a 75 c 111 g 175 t  
 ORIGIN

Query Match 4.4%; Score 171.4; DB 143; Length 533;  
 Best Local Similarity 73.1%; Pred. No. 1.9e-22;  
 Matches 220; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

Qy 498 ttacagcatatggaatagatagatgaactgttttgagagacacagatcgtatcag 557  
 Db 233 TTCTAGATATGGAAATTTTCAACTCAAAATTTGTTGGAGGCCAACAAATAGTTTCAAG 292

Qy 558 ctgatgctgacttaataagatcatatatacaaaacgaaggaagcctcttgaaatgagt 617  
 Db 293 CAGATCAGAGGGCTAAACACATATCTCTCAGATATAAGGACATTAATTTGATTAAT 352  
 Qy 618 atcctagaagatagtggtggtatctcttggaagaatgctgcagatgcttgcttgatgaca 677  
 Db 353 ATCCAGAGATAGTATGATGGATCTCTGTAATAGTCTATGTTAGTTCAAGTTGACAAA 412  
 Qy 678 tgcataagatagagaaatctcgttaactcttaagtcacgacgctcttaagacta 737  
 Db 413 TGCATAGATATAGAGATGATCTCTGTAATTTTGAAGCAATGCTAGGCTAAGGAATC 472  
 Qy 738 tctacttaagatgcttgagagacatcttgcttgcttgatcttgcttgacacaaact 797  
 Db 473 AACTTTAAGTGAAGTTGAAGACATATCTTGTCTTGTGCTCTGTAAGCAAGATT 532  
 Qy 798 c 798  
 Db 533 c 533

RESULT 5  
 LOCUS BG316131 522 bp mRNA EST 26-FEB-2001  
 DEFINITION BG316131  
 ID: Gm-c1032-2967 5' similar to TR:064989 064989 STERIOD  
 22-ALPHA-HYDROXYLASE. [1] ;, mRNA sequence.  
 ACCESSION BG316131  
 VERSION BG316131.1 GI:13125561  
 KEYWORDS EST.  
 SOURCE soybean.  
 ORGANISM Glycine max  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots;  
 Rosidae; euroids I; Fabales; Fabaceae; Papilionoideae; Glycine.  
 REFERENCE 1 (bases 1 to 522)  
 AUTHORS Shoemaker, R., Kelm, P., Vodkin, L., Erpelting, J., Corryell, V., Khanna,  
 A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,  
 Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers,  
 Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurr,  
 R., Rittler, E., Kohn, S., Shiu, T., Jackson, Y., Cardenas, M., McCann,  
 R., Waterston, R. and Wilson, R.  
 TITLE Public Soybean EST Project  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Shoemaker R/Public Soybean EST Project  
 Public Soybean EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available through: Genome Systems, Inc. 4633 World  
 Parkway Circle St. Louis, Missouri 63134 For further information  
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)  
 427-3324 or contact: clones@genomesystems.com or  
 info@genomesystems.com web site: www.genomesystems.com  
 High quality sequence stop: 388.  
 FEATURES  
 source 1..522  
 Location/Qualifiers  
 /organism="Glycine max"  
 /db\_xref="taxon:3847"  
 /clone\_lib="GENOME SYSTEMS CLONE ID: Gm-c1032-2967"  
 /clone\_lib="Gm-c1032"  
 /tissue\_type="Cotyledons of 8-day-old 'Williams' seedlings"  
 /lab\_host="DH10B"  
 /note="Vector: pBluescript.II XR; Site\_1: EcoRI; Site\_2:  
 XhoI; This cDNA library was constructed from mRNA isolated  
 from cotyledons of 8-day-old 'Williams' seedlings which  
 were propagated on paper towels with distilled water for  
 3 days (etiolated), then greenhouse grown for 5 days in  
 potting soil. The cotyledons were flash-frozen in liquid



```

FEATURES
source
    Fax: 864 656 4293
    Email: rwing@clmson.edu
    Seq primer: GCAACACGCTATACCATG
    Class: BAC ends
    High quality sequence stop: 191
    Location/Qualifiers
        1..589

```

/note="Vector: pBel0AC11; Site1: HindIII; Site2: HindIII". Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Armuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from *Oryza sativa*, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9%. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

Query Match	3.8%	Score 146:	DB 233:	Length 589:
Best Local Similarity	57.9%	Pred. No. 1,1e-17:		
Matches 288:	Conservative	0:	Mismatches 191:	Indels 18:
				Gaps 1
QY	500	tacaagtatggtgaatatatagatcogaactgttttggaaaccaacgaatcgtatcagct	559	
Db	85	TGCAGGTACGGGAATATATACCGGTGAGCCCTGTTGCGGGAGACGGCGGTGGTGTGCGGG	144	
QY	560	gatgtctgaactaatagattcatattacaaacgaaggaagcctcttbaatgtagttt	619	
Db	145	GACGGGGGGCTTAACCGGTACATCTCTGCAGAACGAGGGAGGCTGTTCAGTGCACCTAC	204	
QY	620	ectaaagatatagttgttgatctcttggaaatggtcgaatgctgttcttcttgttgtaatg	679	
Db	205	CCGGCCACATCTGGGGGCACTCTGGGCAAGTGGTCCATCTGTCCTGTCGGGGAGCCGG	264	
QY	680	catagaagatatagaagatctcgtttaactcttaagtcgaagcaagcttttagaactatt	739	
Db	265	CACCGCGAATCGGGCCATCTCTCCATCTTCTCTCTCGGTCCGCTCCGCGCGCTGC	324	
QY	740	ctacttaagaatglttgagagacatacttgtttgtcttgatctcttggcaacaaactct	799	
Db	325	CTCCTCCGCCGAGGTGAGAGGCCACACCCCTCTCTGTCCTCGGGCGCTGCCCTTCTCTCC	384	
QY	800	atttctctctgtccaagaagagccaaaaagt-----tttatattt	841	
Db	385	ACCTTCTTGTGCTACACACCAAGCAAGAAAGGTACAAACACACATCGTCTCCATTTAAGTGG	444	
QY	842	atcttttatatttgccaaaatttttctgtttatgaactcttagagtctctaactttttt	901	
Db	445	AGCATGCAATATATCAATGAGCTCGCTCTGATCATATTCATATGMAAGCGCGGTTTGGG	504	
QY	902	tttaattgaacagtttaagtttaacttaactaagtcgaagacataaagatgatatgaccttg	961	
Db	505	TGCGTGGCGGATGTTACAGCTTCAACCTGTATGGCGAAGAACATTAATGATGATGACCCCGG	564	

QY 962 agaagaagaacagagc 978  
| | | | | | | | | |  
Db 565 GGGAGAAGAGACGGAGC 581

RESULT	8
BG043191	

DEFINITION	st98a12.y1	Gm-cl066	Glycine max	CDNA clone	GENOME SYSTEMS	CLONE ID
Gene ID: 10634	10634	10634	10634	10634	10634	10634

22-ALPHA-HYDROXYLASE. [1] ;, mRNA sequence.  
BC043191

VERSION	BG043191.1	GI:12489670
KEYWORDS	EST	

**SOURCE**

ORGANISM

## REFERENCE

## AUTHORS

**AUTHORS**  
Shoenmaker, R., Kelm, P., Voding, L., Erpelund, J., Coryell, V., Khanna,  
A., Bolla, B., Marra, M., Hillier, L., Kuuba, T., Martin, J., Beck, C.,  
Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers,  
Y., Ritten, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurr,  
R., Ritter, E., Kohn, S., Shinn, T., Jackson, Y., Cardenas, M., McCann,  
R., Waterston, R. and Wilson, R.

JOURNAL  
COMMENT

COMMENT:

**COMMENT**  
Contact: Shoemaker R/Public Soybean EST Project  
Public Soybean EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available through: Genome Systems, Inc. 4633 World  
Parkway Circle St. Louis, Missouri 63134 for further information  
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)  
427-3324 or contact: clones@genomesystems.com or  
info@genomesystems.com web site: [www.genomesystems.com](http://www.genomesystems.com)  
High quality sequence stop: 415.

**FEATURES**  
**SOURCE**

1.465

```
/db_xref="taxon:3847"
```

```
/clone_lib="Gm-cl066"
```

old seedling"

```
/note="Vector: pBluescr
```

from unexpanded leaves

seedlings were salt stressed from the cutting

synthesized from mRNA 3 days prior to mat-

poly(U) sequence with adapters were ligated to

followed by  $\lambda$ 101 digest

transformed into DH10B

was constructed in the  
Shoemaker."

BASE COUNT  
ORIGIN

110 C 8/9

Query Match	2.7%	Score	105.4	DB	173	Length	465
Best Local Similarity	74.3%	Pred.	No. 5e-10				
Matches 133		Conservative	0	Mismatches	46	Indels	0
						Gaps	0

Oy	2917	aacacatcatccgcttgtagagaggagccaagggcattgcgttcgttacaagcagcaag	2976
Dd	92	AMCACTTTTCTTCGCTGGAGAGAGACCATATTATGTGCAGAGATCAAGTTAAGTAAG	151
Oy	2977	ttagaatgacagtgctttattcatcatcagtcttaattcaatltggaattagaacaga	3036
Dd	152	CTTGAAATGGCTGTTTTTCATTCACCATCTCATTCCACTACACATTTGGGAATGGCTGAT	211
Oy	3037	gatgataaacaccttgcttcccttcttgttatgttcttctaagcggttgcctattaggct	3095
Dd	212	ACGGATCAAGAAGCTTTTGGCTTACCCCTTTTGTGCGACTTCCCACAAAGGCCCTACCATTTAGAT	270
<b>RESULT 9</b>			
LOCUS	AM398669	615 bp mRNA EST 07-FEB-2000	
DEFINITION	EST309169 L. pennellii trichome, Cornell University Lycopersicon		
ACCESSION	pennelli1 cDNA clone clPTK422 5', mRNA sequence.		
VERSION	AM398669		
KEYWORDS	EST.		
SOURCE	AM398669.1 GI:6917139		
ORGANISM	Lycopersicon pennellii. Lycopersicon pennellii Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Angiosperms; Magnoliopsida; Asteridae: euasterids I; Solanales; Solanaceae; Solanum;		
REFERENCE	Lycopersicon. 1 (bases 1 to 615)		
AUTHORS	Alcala,J., Vrebalov,J., White,R., Matern,A.L., Lakey,J., Holt,I.E., Liang,F., Hansen,T.S., Upton,J., Romling,C.M., Craven,M.B., Fujii, C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin G.B., Tanksley,S.D. and Giovannoni J. Generation of ESTs from wild tomato (Lycopersicon pennellii) trichomes		
TITLE	Unpublished (1999)		
JOURNAL	Contact: David Frisch		
COMMENT	Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Tel: 864 656 4366 Fax: 864 656 4293 Email: dirisch@CLEMSON.EDU 5 prime sequence.		
<b>FEATURES</b>			
source	Location/Qualifiers		
	1..615		
	/organism="Lycopersicon pennellii"		
	/db_xref="taxon:28526"		
	/clone="clPTK422"		
	/clone_1fb-"L. pennellii trichome, Cornell University"		
	/tissue_type="trichome"		
	/dev_stage="mixed stages"		
	/lab_host="SOLR"		
	/note="Vector: pBluescript SK-, Site_1: EcoRI; Site_2: XhoI; Leaves of various stages were shaken in liquid nitrogen, shearing off trichomes. This procedure yielded a mixture of cells highly enriched for trichomes, with minor contamination by other types of leaf cells."		
<b>BASE COUNT</b>			
ORIGIN	180 a 111 c 100 g 224 t		
<b>Query Match</b>			
Best Local Similarity 57.0%; Pred. No.6.1e-10;			
Matches 192; Conservative 0; Mismatches 145; Indels 0; Gaps 0;			
Oy	497	tttacaggatggtgaagatatatagatgaacactgtttgsgaaccaacgacgtatca	556
Dd	265	TGTTCTAGTAGTGGAGAAAGTGTTCAAGTCACATTTATTCTCCCAACACAGTGAGTCA	324
Oy	557	gctgtagtgcgtgaactaatagatctatctatcaaaacgaagaagcctcttgaagtgt	616
Dd	325	TGTGACCAAGACCTTAATTACTTCAATTATTAACAATAAGATTAAGTTATTCAGTGA	384

Y	617	taictagaaagataggggagctcttgggaatgtagatgctgtctgttggtac	676
Db	365	TATCCAAAGCCAAATTCATGATGTTACTTGTGGCAAGTTTCATTGCTGTGGCTGTGGTAC	444
Oy	677	atgcataagatalatagagaatgatactcgttaactcttaagtcgaacgacgcttgaact	736
Db	445	ACACATAAAGAGCTTAGGATGTTTCATTATTCACATATCAGACACATTATAGCTTAACCT	504
Oy	737	attacttaaaagatgttgagagacatacttgttctgttcttgatcttggacaacaaac	796
Db	505	GAGTTATTATTAATGATGTTGCAAACTTACACACTTCAGATTCATTCATCATGAAAGATATAA	564
Oy	797	tctattctctgtcctcaagcagggccaaaaggttt	833
Db	565	CATCAAGTCACATGACTGGGAAGAGCCAGAAAGATT	601
RESULT	10		
BF052051			
LOCUS	BF052051	191 bp	mRNA
DEFINITION	EST437298	tomato clone/immature green fruit	Lycopersicon
ACCESSION	BF052051	esculentum cDNA clone CLEM25M4 5' sequence, mRNA sequence.	
VERSION	BF052051	GI:10805947	
KEYWORDS	EST.		
SOURCE	tomato.		
ORGANISM	Lycopersicon esculentum		
REFERENCE	Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E., Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Rønning,C.M., Niernman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.		
AUTHORS	Generation of ESTs from tomato fruit tissue, Immature green		
TITLE	Unpublished (2000)		
JOURNAL	Contact: David Frisch		
COMMENT	Clemson University Genomics Institute		
	Clemson University		
	100 Jordan Hall, Clemson, SC 29634, USA		
	Tel: 864 656 4366		
	Fax: 864 656 4293		
	Email: dfrisch@CLEMSON.EDU.		
FEATURES	Location/Qualifiers		
Source	1..191		
	/organism="Lycopersicon esculentum"		
	/cultivar="TA496"		
	/db_xref="taxon:4081"		
	/clone="CLEM25M4"		
	/clone_lib="tomato developing/Immature green fruit"		
	/lssue_type="Fruit"		
	/dev_stage="Immature green (5-35 days post-anthesis)"		
	/lab_host="SOLR"		
	/note="Vector: pBluescriptSKMcuadapt; Site:1: EcorI; Site:2: XhoI; Fruit were harvested at 5 dpa (0.5 cm) and harvested at 7 day intervals through 35 dpa. Equal masses of tissue from each stage were combined (including seeds and locules) prior to mRNA isolation."		
BASE COUNT	65 a	19 c	46 g
ORIGIN			61 t
Query Match	2.7%	Score 103.6;	DB 143; Length 191;
Best Local Similarity	71.6%	Pred. NO. 1.1e-09;	
Matches 136; Conservative	0;	Mismatches 54;	Indels 0;
Gaps	0;		
Oy	575	agattcatatcaaaaagagaaggtcttctaattagcttctcctagaagatagct	634
Db	2	AGATACATCTTTCGCAAGAAAGAGAGATTATTGCTGCTAATTATTCACAAAGATATAGCT	61



QY	617	taactcgaagatagaagttggagatctcttggagaatagtcgaatgcttcttcttcttggtagc	676
Db	131	TATCCAAAGCCCAATTCATGGTATATCTGGCAAAAGTTTCATTCGTTGCGGTGGCGAC	190
QY	677	atgcatagagatagagaagtatctgccttaactcttcaagtcagcagctctagaact	736
Db	191	ACACATTAAGAGCGTTTAGSANTGTTTTCATTATCACATATACGACCACTTAAGTCTAAACCT	250
QY	737	attctacttaagaatgttggagagacacttgttctgtcttgatcttgcacacaacac	796
Db	251	GAGTTTATTTATGATGCTTGGAAACATTAGCACTTCAGATTCCTCCAAATCATGGAAGATATA	310
QY	797	tctattctctctgctcaagcagcgcccaaaaagttt	833
Db	311	CATCAAGTCAGATCTCTGGGAAGGCGCAAGAAGCTTTT	347

RESULT	13
LOCUS	BC448552
DEFINITION	BC448552 677 bp mRNA
ACCESSION	NF036B09RT1f1074 Developing root Medicago truncatula cDNA clone
VERSION	NF036B09RT 5', mRNA sequence.
KEYWORDS	BC448552
SOURCE	BC448552.1 GI:1336733
ORGANISM	EST.
	barrel medic.
	Medicago truncatula
	16-MAR-2001

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots; Rosidae: eucotsids I: Fabales; Fabaceae; Papilionoideae; Medicago. '1 (bases 1 to 677)	Watson, B.S., Shuh, H.-S., Lopez-Meyer, M., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J., Flores, H.R., Imman, J.T., Weller, J.W., May, G.D. and Palva, N.L.	Expressed Sequence Tags from the Samuel Roberts Noble Foundation Medicago truncatula root library	Unpublished (2000)	Contact: Palva NL

Plant Biology Division  
The Samuel Roberts Noble Foundation  
2510 Sam Noble Parkway, Ardmore, OK 73402, USA  
Tel: 580 221 7317  
Fax: 580 221 7380  
Email: [nlp@srnoble.org](mailto:nlp@srnoble.org)  
Insert Length: 677 Std Error: 0.00  
Plate: 036 Row: B Column: 09  
Seq primer: TCACCAAGGAACAGACTATGAC.

FEATURES	SOURCE
location/Qualifiers	1. 677
/organism="Medicago truncatula"	
/db_xref="taxon:3880"	
/clone="NF036B09RT"	
/clone_lib="developing root"	
/tissue_type="root"	
/dev_stage="Pooled developmental"	
/note="Vector: lambda Zap; Total RNA was extracted from non-nodulated roots of plants grown in 1 mM nitrate medium. Samples were taken at four time points (approximately two days, one, two and six weeks post germination) representing early seedling growth to nitrogen limitation."	
BASE COUNT	248 a 11 c 135 g 179 t 4 others
ORIGIN	

Query Match 2.6% Score 100 DB 153 Length 677;  
 Best Local Similarity 64.9% Pred. No. 5.3e-09;  
 Matches 148; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

Oy 1268 gaagcagcagatgagtgtaagatgcatcattgttgggaacaagaacagacagatcattcttg 1327  
 357 GAAGCAAGCAGATAAAGATCAACAGCGGAATGAAATTTCGAGAGAAATGATCTTTTGG 416

[illegible]

RESULT		14	
AM690507			
LOCUS	AM690507	524 bp	mRNA
DEFINITION	NC033D01ST1P1000 Developing stem Medicago truncatula cDNA clone	EST	20-DEC-2000
ACCESSION	NM033D01ST_5'		mRNA sequence.
VERSION	AM690507		
KEYWORDS	AM690507.2	GI:11933719	.
SOURCE	EST.		
	barrel medic.		

ORGANISM  
Medicago truncatula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta  
Rosidae; eucotsids I; eudicotyledons; core eucotsids;  
Rosidae; eucotsids I; Fabales; Fabaceae; Papilionoideae; Medicago.  
1 (bases 1 to 524)  
REFERENCE  
He, X.-T., Shadle, G., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell,  
C.J., Flores, H.R., Inman, J.T., Weller, J.W., May, G.D. and Dixon  
R.A.

TITLE	Expressed Sequence Tags from the Samuel Roberts Noble Foundation Medicago truncatula stem library
JOURNAL	Unpublished (2000)
COMMENT	On Apr 14, 2000 this sequence version replaced g1:7555243.

Contact: Dixon RA  
 Plant Biology Division  
 The Samuel Roberts Noble Foundation  
 2510 Sam Noble Parkway, Ardmore, OK 73402, USA  
 Tel: 580 221 7302  
 Fax: 580 221 7380  
 Email: radixon@noble.org  
 Insert Length: 670 Std Error: 0.00  
 Plate: 035 row: D column: 01  
 Seq primer: TCACACAGAAACAGCATGAC.  
 Location/Qualifiers

```
source
1. 524
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="MF035D01SR"
/clone_lib="Developing stem"
/tissue_type="stem"
/dev_stage="Pooled developmental"
/Note="Vector: Lambda Zap; Contains a mixture of
internodal stem segments"
BASE COUNT 172 a 90 c 86 g 176 t
ORIGIN
```

Query Match	2.5%	Score 97.2	DB 119	Length 524
Best Local Similarity	79.6%	Pred. No. 1.8e-08		
Matches 164	Conservative	0	Mismatches 38	Indels 4
				Gaps 4
QY 503	aggtatgttaagatatacagatcgacctgcttttggagaacaacagatcgtatcagtcgat	562		
Db 319	AGGTAATGGTAAATTTTACAAATCAAACTTTGTTGGTGGACCAACAATAGTTGCACACAT	378		
QY 563	gcttgagact-aaatagatcatalatcaaaacaggaagagcgtccttgaag-aaagttatc	620		
Db 379	GCAGGATGGAATAGGTCATATTTCACAAATGAAGGAAATATTATTTGATGCAATTTATC	438		
QY 621	ctagaagatcatagtggtggagatccttgggaaatggtcgatcgtcgtcttc-gttgggtaca-t	678		
Db 439	CTAGTAGCATTTGGTGGATTAAGTGGGAAATGGCTCATTTGGTTTGGTTGGTGCACATTT	498		

QY	679	gcata	gagata	tatgaga	agtatctgc	704
Db	499	GCATAGAGACATGCGTAA	TATATC	CAC	524	

RESULT	15
AM759817	
LOCUS	
DEFINITION	AM759817 461 bp mRNA EST 18-JUL-2000
	S154d10.y1 Gm-cl027 Glycine max CDNA clone GENOME SYSTEMS CLONE ID

ACCESSION	AW759817	
VERSION	AW759817.1	GI:7691694
KEYWORDS	EST.	
SOURCE	soybean.	
ORGANISM	Glycine max	

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 461)	Shoemaker, R., Keim, P., Vodka, L., Erpelting, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kueba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepleco, M., Theising, B., Allen, M., Bowers, T., Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurr, R., Ratter, E., Kohn, S., Shu, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.	Public Soybean EST Project	Unpublished (1999)	Contact: Shoemaker R/Public Soybean EST Project

FEATURES	Location/Qualifiers
source	1. .461

constructs were size-fractionated using GibcoBRL Life Technologies' cDNA Size Fractionation column. The column eluent was then ligated into StrataGene's pBluescript(1m) II XR Predigested Vector (pBluescript II SK(+)) that has been digested with EcoRI and XhoI, and phosphorylated by StrataGene). 97% of the white and blue colonies appear to contain recombinant plasmids with cDNA inserts, based on size (n=30). This library was constructed by Dr. Paul Keim and Dr. Virginia Corryell."

Search completed: October 17, 2001, 16:25:13  
Job time: 6273 sec

ORIGIN	BASE COUNT	149 a	90 c	109 g	113 t
Query Match	2.5%:				
Best Local Similarity	76.1%:				
Matches 118:	Conservative	0:	Mismatches 37:	Indels 0:	Gaps 0:
Oy	913	agttacgcttaatcctaataggaacataataatgaatgatgccttggaagaagaana	972		
Db	99	AGTTACCTTCATTTGATGGCCAGCATATCATGAGCATGATGCATCCAGGGATATAGAGA	158		
Oy	973	cagagcaatlaaagaagaagatgtaacttcaatgaagaagagtgtctctgctcctaa	1032		
Db	159	CAGAGCATCTTAAAGAGAGAGTACGTACCTTTTCATGAAGGGGTGTATCCGCCGTCATTAA	218		
Oy	1033	atctaccagaagactgcttatcataaagccttcacg	1067		
Db	219	ATTTACCTGGAACTGCATATCCGAAAGGCATTGAG	253		





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2001, 19:48:49 ; Search time 9769.47 Seconds  
(without alignments)  
566.812 Million cell updates/sec

Title: US-09-502-426-1\_COPY\_6111\_6468  
Perfect score: 358  
Sequence: 1 aaaaaaaaaagatgaaagt.....tgcataataccacaacag 358

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 773874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_ba3:\*  
4: gb\_in1:\*  
5: gb\_in2:\*  
6: gb\_in3:\*  
7: gb\_om:\*  
8: gb\_ov:\*  
9: gb\_pat1:\*  
10: gb\_pat2:\*  
11: gb\_ph:\*  
12: gb\_pl1:\*  
13: gb\_pl2:\*  
14: gb\_pl3:\*  
15: gb\_pl4:\*  
16: em\_ba1:\*  
17: em\_ba2:\*  
18: em\_fun:\*  
19: em\_htgo\_hum:\*  
20: em\_htgo\_inv:\*  
21: em\_htgo\_rod:\*  
22: em\_htgo\_hum1:\*  
23: em\_htgo\_hum2:\*  
24: em\_htgo\_hum3:\*  
25: em\_htgo\_hum4:\*  
26: em\_htgo\_hum5:\*  
27: em\_htgo\_hum6:\*  
28: em\_htgo\_hum7:\*  
29: em\_htgo\_hum8:\*  
30: em\_htg\_inv1:\*  
31: em\_htg\_inv2:\*  
32: em\_htg\_other:\*  
33: em\_htg\_rod:\*  
34: em\_hum1:\*  
35: em\_hum2:\*  
36: em\_hum3:\*  
37: em\_hum4:\*  
38: em\_hum5:\*  
39: em\_hum6:\*  
40: em\_hum7:\*  
41: em\_in:\*  
42: em\_om:\*  
43: em\_or:\*

44: em\_ov:\*  
45: em\_pat:\*  
46: em\_ph:\*  
47: em\_pl:\*  
48: em\_ro:\*  
49: em\_sts:\*  
50: em\_sy:\*  
51: em\_un:\*  
52: em\_v1:\*  
53: gb\_sts1:\*  
54: gb\_sts2:\*  
55: gb\_sts3:\*  
56: gb\_sy:\*  
57: gb\_un:\*  
58: gb\_v11:\*  
59: gb\_v12:\*  
60: gb\_htg1:\*  
61: gb\_htg2:\*  
62: gb\_htg3:\*  
63: gb\_htg4:\*  
64: gb\_htg5:\*  
65: gb\_htg6:\*  
66: gb\_htg7:\*  
67: gb\_htg8:\*  
68: gb\_htg9:\*  
69: gb\_htg10:\*  
70: gb\_htg11:\*  
71: gb\_htg12:\*  
72: gb\_htg13:\*  
73: gb\_htg14:\*  
74: gb\_htg15:\*  
75: gb\_htg16:\*  
76: gb\_htg17:\*  
77: gb\_htg18:\*  
78: gb\_htg19:\*  
79: gb\_htg20:\*  
80: gb\_htg21:\*  
81: gb\_htg22:\*  
82: gb\_htg23:\*  
83: gb\_htg24:\*  
84: gb\_htg25:\*  
85: gb\_pr1:\*  
86: gb\_pr2:\*  
87: gb\_pr3:\*  
88: gb\_pr4:\*  
89: gb\_pr5:\*  
90: gb\_pr6:\*  
91: gb\_pr7:\*  
92: gb\_pr8:\*  
93: gb\_pr9:\*  
94: gb\_ro1:\*  
95: gb\_ro2:\*  
96: gb\_in4:\*  
97: gb\_pr10:\*  
98: em\_ba3:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	358	100.0	4818	12	AF044216 Arabidops
2	356.4	99.6	84196	14	ATT3A5 Arabidops
3	56.6	15.8	1147	53	CNS073CX Arabidops
4	55.6	15.5	169546	60	AC004157 Plasmodu
5	55	15.4	1007	53	CNS06X9S Arabidops
6	54.2	15.1	204652	84	PFMAL13P6 Arabidops
7	53.6	15.0	1167	53	CNS07360 Arabidops
8	53.4	14.9	973	53	CNS071LE Arabidops



Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 84196)  
Bloembergen, H., Mewes, H.W., Lemcke, K., Mayer, K.F.X., Quetier, F. and Salanoubat, M.  
JOURNAL  
Unpublished  
REFERENCE  
2 (bases 1 to 84196)  
EU Arabidopsis sequencing project.  
AUTHORS  
Direct Submission  
TITLE  
Submitted (25-JAN-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de, project Coordinator: Marcel Salanoubat and Francis Quetier, Groupement d'Interet Public, Centre National de Sequencage - GENOSCOPE, 2 rue Gaston Creneau, BP191, 91006 Evry Cedex, France; http://www.genoscope.cns.fr  
COMMENT  
On Jan 27, 2000 this sequence version replaced gi:6434247. Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.  
FEATURES  
source  
1. 84196  
/organism="Arabidopsis thaliana"  
/variety="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="3"  
2983. .3338  
/note="355 bp LTR"  
3339. .7564  
/note="LTR-transposon"  
7565. .7922  
/note="367bp LTR"  
12339. .13197  
/gene="T3A5.10"  
join(12339. .12665,12742. .12803,12888. .13055,13125. .13197)  
/gene="T3A5.10"  
/note="Similarity to cyclin-dependent kinase inhibitor (ICK1), Arabidopsis thaliana, EMBL:AT094772.1"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAB62432.1"  
/db\_xref="GI:561966"  
/translation="MAAARRRRRRVVEENGVTYYVKKRKKKEEVDLVESRIILSPCY QATNRGIVARNSAGASVIVRRDPSVPEOCIEEDSSVCCSTSEKSKRR IEFYDLNNGDDRETSMTYDLNKSSEMMNDSVAVEDVESRRRLKSLHEVY KEALDFEFOYAEEDLNKLKLECSMKYNFDEKDEPLOGGRYEMVKLPN"  
12339. .12665  
/gene="T3A5.10"  
/number=1  
12666. .12741  
/gene="T3A5.10"  
/number=1  
12742. .12803  
/gene="T3A5.10"  
/number=1  
12804. .12887  
/gene="T3A5.10"  
/number=2  
12888. .13055  
/gene="T3A5.10"  
/number=3  
13056. .13124  
/gene="T3A5.10"  
/number=3  
13125. .13197  
/gene="T3A5.10"  
/number=4  
complement(15858. .15901)  
/number=1  
15858. .16470  
/gene="T3A5.20"  
complement(join(15858. .15901,16014. .16470))  
/note="similarity to predicted proteins, Arabidopsis

thaliana"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAB62433.1"  
/db\_xref="GI:561967"  
/translation="MARLEMHNSNGARISFNSFEVIREKSNKSNNSNFSF MSPADFAFVYDYSMTIPADEFLGLKILPEKTSVHVRITGLCEELLTEEGSMVGNTF SLRPLILSSSFSTKGTWRELLGLKRTIHRSKKTKDKVNEEVLSDHKIISGNVATRC OVADTR"  
complement(15902. .16013)  
/number=1  
complement(16014. .16470)  
/number=2  
18303. .19931  
/gene="T3A5.30"  
complement(18303. .19931)  
/number=1  
complement(18303. .19931)  
/codon\_start=1  
/product="scarecrow-like 7 (SCL7)"  
/protein\_id="CAB62434.1"  
/db\_xref="GI:561968"  
/translation="MAYTCIDSGNLMALIAQILIKOKQOQSQHQOEEQDEPNPWN PSFGFTLPSCGFSDPFQVNDGPHFPHLEHQAASAEEFSDDEMSELINGDAS QTNDDPFIYGHDPFSPFSLAPSVLNVRKDSASQOLPPPSAPSPSPQ HPPEPPQPDNLNPIFKATHDYARKPEPDPDLIRIKESVSGDPIORGVYFAE ALSHKPEPSSSSSSLEDFILSYKTLNDCPSKFAHLITANALITATNOSNIHI VDFGIFQIGMSALLOALATRSCKPRTIRISGIPAPSLGDSPPSLATGNRLRDRA AILDNFEFYVLTPIQLNGSSFRVDPDEVLVNFMELKTLDETATYVGLRLA RSLNPRIVTLEGEVSLNRYEAFNRKNSLRFSAVESLEPNLDRSKERLYERVL FGRRIMDLVRSDDDNKRGTFGLMEEEKQQRVMEKAGFEPVPRSNVAASOAKLLIM NYNSTLYSLVSESEPGFISLAWNVPLTVSSWR"  
26093. .28999  
/gene="T3A5.40"  
complement(join(26093. .26338,27091. .27200,27308. .27386, 27478. .27570,27723. .28037,28136. .28288,28374. .28698, 28779. .28999))  
/note="Contains cytochrome P450 cysteine heme-iron ligand signature AA455-464: prokaryotic membrane lipoprotein lipid attachment site AA320-330"  
/codon\_start=1  
/product="steroid 22-alpha-hydroxylase (DWF4)"  
/protein\_id="CAB62435.1"  
/db\_xref="GI:561969"  
/translation="MPEFEHHTLPLILLPSLSILLFLILKKRRNRTRNLPGRKS GMPLIGETIGLAKYTRATTLGDDPMQOHRSKYKGYRSLREEPITYSDAGLNPTILO NEGRLEFCSTYRSTIGTIGKWSMLVGDMDRMSISLNTLSHRLKTLIKDVERH TLFVLDWMQONSIFSAQDEARKFTFNLMARKIMSDPEEETEDLKEVYFMGVSV APLNLPAYAKHALQSRATILKFTERKKEERKLDIKEDQEEVEVTEDEAKMSOH VRKORTDDLLGWLKHSNSTEOILDLITLSLPAHGHTSSVAALAIFFIQAQPKAV ELREBEHLERAKKELGESELMNDQVKKMFTQCIVINETRLGNVAFHLRKLAKOV RYKGYDIPSGKXVLPVISAHLVLDNSRXPQPLFPMRMOQONNKAASSGSGSPSTMN NYMPFGGPRILCAGSELAKLEMAVFIHHLVLKFWELAEEDKPAFPVDPVDPNLPILR VSRIL"  
complement(26093. .26338)  
/number=1  
complement(26339. .27090)  
/number=1  
complement(27091. .27200)  
/number=2  
complement(27201. .27307)  
/number=2  
complement(27308. .27386)  
/number=3  
complement(27387. .27477)  
/number=3  
complement(27478. .27570)  
/number=4  
complement(27571. .27722)  
/number=4  
complement(27723. .28037)  
/number=5  
complement(28038. .28135)  
introns

```

exon      /number=5
           /complement(28136..28288)
intron    /number=6
           /complement(28289..28373)
exon      /number=6
           /complement(28374..28698)
intron    /number=7
           /complement(28699..28778)
exon      /number=7
           /complement(28779..28999)
gene       /number=8
           /38307..41323
           /gene="T3A5.50"
exon      /complement(38307..38796)
           /number=1
           /complement(join(38307..38796,38877..39090,39189..39270,
40180..40242,40337..40401,40489..40564,40655..40734,
40805..40919,41225..41323))
           /note="Contains Eukaryotic putative RNA-binding region
RNP-1 signature AA179-186"
           /codon_start=1
           /product="rpl snrnp 70k protein"
           /protein_id="CA862436.1"
           /db_xref="GI:6561970"
           /translation="MGDSGDFLRNPNAAVQARAKVONRANVLQKLMGSHPTGLTN
NLKLPFRPPELEKPPPEKPCPTGMAQFVNFAPGPEYAPAPKPEVLEPSOKR
ERIKRLKLEGEKAEADLKYPNDNDPNTGDEYKTLFSLRVESSEKIKREPS
YGPRIKRVHLYTDLTNPKGAFIEVMTDMKAAYKQADGQKIDGRRVLDVGRGT
VPMNRPRLGGGLGTSRYVGEQELVGEQDQGRISQSEPSRPREKREKREKERE
RSRLSHQPEKRSRDRPREKDHNRDQGRDSDRSRDRSRDRSGDRDGRDRDR
GRDRTSRDHDNRKRKREKERYEGEVEHGGGRSDRSRDRSGDRDGRDRDR
TDRYSRYDKMEEDDFRYEREYKRSKRSSEKREYVR"
intron    /complement(38797..38876)
           /number=1
exon      /complement(38877..39090)
           /number=2
intron    /complement(39091..39188)
           /number=2
exon      /complement(39189..39270)
           /number=3
intron    /complement(39271..40179)
           /number=3
exon      /complement(40180..40242)
           /number=4
intron    /complement(40243..40336)
           /number=4

Query Match      99.6%: Score 356.4; DB 14; Length 84196;
Best Local Similarity 99.7%: Pred. No. 3.8e-49;
Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Db 25792 AACTTTACCTGCTTTTGGATCGAAGCAATTAACAATTTGTCAATACCAACACACAG 25735
RESULT      3
CNS073CX    1147 bp  DNA      STS      11-JAN-2001
LOCUS       CNS073CX
DEFINITION  clone BA0AB018F06 of library BA0AB from strain CLIB 210 of
            Kluyveromyces lactis, sequence tagged site.
ACCESSION   AL427351
VERSION     AL427351.1  GI:12210545
KEYWORDS    STS.
SOURCE      Kluyveromyces lactis.
            Kluyveromyces lactis.
ORGANISM    Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
REFERENCE   1 (bases 1 to 1147)
AUTHORS     Bolotin-Fukuhara,M., Toffano-Nioche,C., Artiguenave,F.,
            Duchateau-Nguyen,G., Lemaire,M., Matheisse,R., Montrocher,R.,
            Robert,C., Termier,M., Winkler,P. and Wesolowski-Louvel,M.
            Genomic Exploration of the Hemiascomycetous Yeasts: 11.
            Kluyveromyces lactis
            FEBS Lett. 487 (1), 66-70 (2000)
TITLE       1152886
JOURNAL     PUBMED
REFERENCE   2 (bases 1 to 1147)
AUTHORS     Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brothier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuvéglise,C., Ozler-Kalogeropoulos,O., Potier,S.,
            Saurin,M., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Winkler,P. and Weissenbach,J.
            Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
            yeast species for molecular evolution studies(1)
            FEBS Lett. 487 (1), 3-12 (2000)
TITLE       1152876
JOURNAL     PUBMED
REFERENCE   3 (bases 1 to 1147)
AUTHORS     Direct Submission
            Submitted (08-SEP-2000) Genoscope - Centre National de Sequençage,
            2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            segre@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This STS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,
            saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES
    source
        1..1147
        /organism="Kluyveromyces lactis"
        /strain="CLIB 210"
        /variety="lactis"
        /db_xref="taxon:28985"
        /clone="BA0AB018F06"
        /clone_1bp="BA0AB"
    STS
        1..1147
        /note="part of mitochondrial DNA"
        /evidence=noe-experimental
    BASE COUNT
        284 a 120 c 175 g 361 t 207 others
    ORIGIN
        Query Match      15.8%: Score 56.6; DB 53; Length 1147;
        Best Local Similarity 39.6%: Pred. NO. 1.6;
        Matches 90; Conservative 39; Mismatches 98; Indels 0; Gaps 0;

```

OY	70	ccaatgcatataaaaaatttggaacaaatatcatatcgctttttagtcgg	129
Dd	745	TTTTTTT TTTT TTTATATATWAAAAAAAMTWAAAAGATAATTTTTTTTKTKGGGR	804
OY	130	gttcgagaagaagggttcgcacttcgaaaagtgagcagtatalatgatggagcctaqt	189
Dd	805	RATMAAANAAGAWAKAGTDMMWAANAATAAGRTTWWMTAAATATKMWDTGCGKT	864
OY	190	tgaagcttggacaattgtaattgatgttgttatattaatgatgctga	236
Dd	865	TTTTTTTTTTTAATKTKGGGGRRTTTTKKKKKKGKKAAGA 911	
<b>RESULT 4</b>			
AC004157	LOCUS	AC004157 169546 bp DNA HTG	12-AUG-2000
DEFINITION		Plasmodium falciparum chromosome 12 clone 3D7, ** SEQUENCING IN PROGRESS **	2 unordered pieces.
AC004157	ACCESSION	AC004157	
KEYWORDS		HTG; HTGS; PHASE1.	
SOURCE		malaria parasite P. falciparum.	
ORGANISM		Plasmodium falciparum	
REFERENCE		Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
AUTHORS		Hyman,R.W., Fung,E.L., Qin,F., Rowley,D., Mao,J., Tanaki,T., Kurdil,O.B., Conway,A.B. and Davis,R.W.	
JOURNAL		Plasmodium falciparum 3D7 chromosome 12 unpublished	
REFERENCE		2 (bases 1 to 169546)	
TITLE		Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.	
AUTHORS		Direct Submission	
JOURNAL		Submitted (19-FEB-1998). Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA	
COMMENT		On Aug 12, 2000 this sequence version replaced gi:8910447.	
		* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.	
		* 1 23466: contig of 23466 bp in length	
		* 23467 23666: gap of unknown length	
		* 23667 169546: contig of 145880 bp in length.	
FEATURES		Location/Qualifiers	
SOURCE		1..169546 /organism="Plasmodium falciparum"	
		/db_xref="taxon:5833"	
		/chromosome="12"	
		/clone="PFYAC293"	
		/clone="3D7"	
BASE COUNT		69871 a 15381 c 15705 g 68389 t 200 others	
ORIGIN			
Query Match		15.5% . Score 55.6: DB 60: Length 169546;	
Best Local Similarity		51.6% . Pred. NO. 1.5; Indels 0; Gaps 0;	
Matches 127: Conservative		0; Mismatches 119; Indels 0; Gaps 0;	
OY	1	aaaaaaaaaagatgaagatattctatccctctcttttttgcataattcaatca	60
Dd	136495	AAAAAAAAAAAAAAAAAGATRAAACATTGAATCTTTTGATGTCTTTACCCTTTT	136554
OY	61	ttttttgccaatgatataaaaaatttggaacaaatatcatatcgatattcg	120
Dd	136555	TTTTTTTTTCTTTTGCTACCAAGAAATAATATTCATATTTTGGAATTTCTTGA	136614
OY	121	ttagtcgggtttgagaagaagggttcgcacttcgaaagtgcagatgatata	180
Dd	136615	TGAATATGGAATTAATAAACAACCTATGCTCATCAATTAATACAATATATCTCCCTTTGG	136674

Oy 136734 GCAAATGAAAAAATATATACATATTATATATAGACATTTATATATATATATATATATATATA 136734  
 Oy 241 attaa 246  
 Db 136735 TATATA 136740  
 RESULT 5  
 CNS06X9S/C  
 LOCUS  
 DEFINITION T3 end of clone AX0AA039F08 of library AX0AA from strain CBS 7064  
 ACCESSION AL419462  
 VERSION AL419462.1 GI:12202640  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Pichia farinosa.  
 Pichia farinosa  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 Saccharomycetales; Saccharomycetaceae; Pichia.  
 REFERENCE 1 (bases 1 to 1007)  
 AUTHORS de Montigny,J., Spehner,C., Souciet,J., Tekala,F., Dujon,B.,  
 Wincker,P., Artiguenave,F. and Potier,S.  
 TITLE Genomic Exploration of the Hemiascomycetous Yeasts: 15. Pichia  
 scotbiphila  
 JOURNAL FEBS Lett. 487 (1), 87-90 (2000)  
 PUBMED 11152890  
 REFERENCE 2 (bases 1 to 1007)  
 AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,  
 Boloitin-Fukuhara,M., Bon,E., Brotier,P., Casaregola,S.,  
 de-Montigny,J., Dujon,B., Durenz,P., Leplinge,A., Lorente,B.,  
 Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,  
 Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Douvet,M.,  
 Wincker,P. and Weissenbach,J.  
 TITLE Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of  
 yeast species for molecular evolution studies(1)  
 JOURNAL FEBS Lett. 487 (1), 3-12 (2000)  
 PUBMED 11152876  
 REFERENCE 3 (bases 1 to 1007)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Séquençage,  
 2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :  
 segrel@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
 COMMENT This STS is part of a random genomic sequencing program of thirteen  
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces  
 exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,  
 Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces  
 lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia  
 angusta, Debaryomyces hansenii var. hansenii, Pichia scotbiphila,  
 Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to  
 5 kb were prepared and both extremities were sequenced. See  
 keywords for description of this sequence and for the sequence of  
 the other extremity of this insert.  
 FEATURES  
 source  
 location/Qualifiers  
 1..1007  
 /organism="Pichia farinosa"  
 /strain="CBS 7064"  
 /db\_xref="taxon:4920"  
 /clone="AX0AA039F08"  
 /clone\_11b="AX0AA"  
 /note="end : T3"  
 1..1007  
 BASE COUNT 533 a 86 c 126 g 161 t 101 others  
 ORIGIN  
 Query Match 15.4% Score 55; DB 53; Length 1007;  
 Best Local Similarity 39.9% Pred. NO. 3;  
 Matches 109; Conservative 37; Mismatches 127; Indels 0; Gaps 0;

[illegible]

RESULT	6
PfMAL13P6	
LOCUS	PfMAL13P6_204652 bp DNA HTG 19-AUG-1999
DEFINITION	Plasmodium falciparum chromosome 13 strain 3D7, *** SEQUENCING INCOMPLETE
ACTION	PROGRAMS ***, In unordered pieces.
VERSION	AL049183.5 GI:5763804
KEYWORDS	HTG; HTGS-PHASE1.
SOURCE	malaria parasite P. falciparum.

ORGANISM *Plasmodium falciparum*  
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
REFERENCE 1 (bases 1 to 204652)  
AUTHORS Bowman, S., Churches, C., Harris, B., Harris, D., Lawson, D., Quail, M.,  
and Barrell, B.  
TITLE Direct Submission  
JOURNAL Submitted (15-MAR-1999) P.falciparum Genome Sequencing Consortium  
The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridg  
CB10 1SA, UK  
COMMENT On Aug 24, 1999 this sequence version replaced gi:5731886.

For more information about this sequence or the Malaria Project see [http://www.sanger.ac.uk/Projects/P\\_falciparum](http://www.sanger.ac.uk/Projects/P_falciparum). IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc.

Order of segments is not known; 800 n's separate segments.

\* NOTE: This is a 'working draft' sequence.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will be preserved.

FEATURES	Location/Qualifiers
source	1. .204652

BASE COUNT	68880 a	17396 c	17710 g	67856 t	32810 others
ORIGIN	/organism="Plasmodium falciparum" /strain="3D7" /db_xref="taxon:5833" /chromosome="13"				
Query Match	15.18;	Score 54.2;	DB 84;	Length 204652;	
Best Local Similarity	51.4%	Pred. No. 2.5;			
Matches 123; Conservative	0;	Mismatches 118;	Indels 0;	Gaps	

Qy 1 aaaaaaaaaagatgaaagtattttctctccttttttttgataatttaacac 60  
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
 Db 93291 ATATTAATAAATAATATATATATATATCTATTTTGATATATATTTTAAATCT 93350  
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
 Qy 61 tttttttgcccaatgatataataaaatttgataaataatattatgtaattcgttt 120  
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Accession	Sequence	Position
Db 93351	ATTATTATTTCCTATATATAATAATTTTGTAAAAATTATTATTAAGAATGTTANA	93410
Oy 121	ttagttcgggtttgagaaagggttcgacttcgaaagtgcagcgtatagattgg	180
Db 93411	ATGAAATTGAGTATAAATCAATCTGAGGAGTTAGTTGATGATTAATATAATATGAATT	93470
Oy 181	gagctaaagttgaagctcttcggacatttgtatcgatgctgtcatcatcagtcgacact	240
Db 93471	TAAAAAACTTTAGGATTTAAGCTTTTTTTTTCGATGATGAAGAAAATAAAACTACATTAT	93530
Oy 241	att 243	
Db 93531	TTT 93533	

RESULT	7	
CNS07360		
LOCUS		
DEFINITION	CNS07360	1167 bp DNA
	clone BA0A08017A09 of library BA0AB from strain CLR8 210 of	STS
	KluYveromyces lactis, sequence tagged site.	
ACCESSION	AL427102	
VERSION	AL427102.1	GI:12210296
KEYWORDS	STS.	
SOURCE	KluYveromyces lactis.	
ORGANISM	KluYveromyces lactis	
	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;	

REFERENCE	AUTHORS	TITLE	JOURNAL	PUBLISHED	REFERENCE
1 (bases 1 to 1167)	Saccharomycetales: Saccharomycetaceae; Kluyveromyces.				
1 (bases 1 to 1167)	Boletín-Fukuhara, M., Toffano-Nioche, C., Artiguenave, F., Duchateau-Nguyen, G., Lemaître, M., Marmelsse, R., Montrocher, R., Robert, C., Termier, M., Wincker, P., and Wesołowski-Jouvet, M.	Genomic Exploration of the Hemiascomycetous yeasts: 11. Kluyveromyces lactis	FEBS Lett. 487 (1), 66-70 (2000)	11152886	
2 (bases 1 to 1167)					

AUTHORS	Souchet, J.-L., Aigle, M., Artiguenave, F., Blandin, G., Boltoum-Puklatura, M., Bon, E., Brotier, P., Casaregola, S., de-Montigny, J., Dujon, B., Durieux, P., Lepingle, A., Liorette, B., Maretty, A., Neugegelle, C., Ozier-Kalogeropoulos, O., Potier, S., Saurin, W., Tekata, F., Toffano-Nicchoe, C., Wesolowski-Louvel, M., Wincker, P. and Weissenbach, J.
TITLE	Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of yeast species for molecular evolution studies(1)
JOURNAL	FEBS Lett. 487 (1), 3-12 (2000)
PUBMED	11152876
REFERENCE	3 (bases 1 to 1167)

AUTHORS	Genoscope, .
TITLE	Direct Submission
JOURNAL	Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage 2 rue Gaston Cremlieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail sebastien@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT	This STS is part of a random genomic sequencing program of thirt

yeast species: *Saccharomyces bayanus* var. *uvorum*, *Saccharomyces exiguus*, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*, *Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces lactis* var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*, *Candida tropicalis* and *Yarrowia lipolytica*. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

/organism="Kluyveromyces fragilis"  
 /strain="CLIB 210"  
 /variety="lactis"  
 /db\_xref="taxon:28985"  
 /clone="BA0AB017A09"  
 /clone\_1lb="BA0AB"  
 1..1167  
 /note="part of mitochondrial DNA"  
 /evidence="not\_experimental"



COMMENT	BASE COUNT
<p>Direct Submission</p> <p>Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : <a href="mailto:seget@genoscope.cns.fr">seget@genoscope.cns.fr</a> - Web : <a href="http://www.genoscope.cns.fr">www.genoscope.cns.fr</a>)</p> <p>This STS is part of a random genomic sequencing program of fifteen yeast species: <i>Saccharomyces bayanus</i> var. <i>uvaurum</i>, <i>Saccharomyces exiguus</i>, <i>Saccharomyces servazzii</i>, <i>Zygosaccharomyces rouxii</i>, <i>Saccharomyces kluyveri</i>, <i>Kluyveromyces thermotolerans</i>, <i>Kluyveromyces lactis</i> var. <i>lactis</i>, <i>Kluyveromyces marxianus</i> var. <i>marxianus</i>, <i>Pichia angusta</i>, <i>Debaryomyces hansenii</i> var. <i>hansenii</i>, <i>Pichia sorbitophila</i>, <i>Candida tropicalis</i> and <i>Yarrowia lipolytica</i>. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.</p> <p>Location/Qualifiers</p> <p>1. .573</p> <p>/organism="Saccharomyces servazzii"</p> <p>/strain="CBS 4311"</p> <p>/db_xref="taxon:27293"</p> <p>/clone="XAT0A001E06"</p> <p>/clone_lib="XAT0A"</p> <p>/note="end : 17"</p> <p>1. .573</p> <p>/note="part of mitochondrial DNA"</p> <p>/evidence="not experimental"</p>	<p>427 a 33 c 14 g 64 t 35 others</p>

Query Match	Similarity	14.8%	Score 53:	DB 53:	Length 573:
Best Local	Similarity	42.8%	Pred. No. 6.7:	Mismatches 148:	Indels 0; Gaps 0;
Matches	128:	Conservative	23:	Mismatches	148;
Qy	20	tattttatctctctcttttttttttttga	taattttaa	tcatttttttgcctccaa	gata 79
Db	351	TATTTTATTTTATGATTTTATTTTATTTT	TTTTTTTTTTTATTTTATTTTGT	TTTTTTTGGATTTT	TTA 292
Qy	80	tataaaatttggataaataatattatcg	gatattc	gatttttagctcggttg	agaaa 139
Db	291	TATTTTATTTTGTGTAGTTTATTTTAT	KATTTTATTTTATTTTATTTTAACTT		232
Qy	140	agggtctgacttcggaagtcgacgata	gatata	gatcttgsgagctagtgagctttg	199
Db	231	ATTTTGTGACGTTTATTTTAAATTTT	ATTTTATTTTATTTTATTTTATTTT	TTTTTATTTTATTTTATTTT	172
Qy	200	gacatttgatctgagtcgttgatattag	ctgcgcac	atataaacttaaatgagct	259
Db	171	GGGCTATTTTATTTTATTTTATTTT	ATTTTATTTTATTTTATTTTATTTT	TTTTTATTTTATTTTATTTT	112
Qy	260	ttccataaggcccaattatatacgatt	ataataca	aagtcgacaactttactcg	tttt 318
Db	111	WTTKTTTTTTTTATTTTATTTTWT	TGTTTWTGTTTMTTATATATTTT	TKMTATATTTATTTATTTAT	53
RESULT	10				
LOCUS	AC005139				
DEFINITION	AC005139	256172 bp	DNA	HTG	01-APR-1999
ACCESSION	AC005139	Plasmidium falciparum	chromosome 12,	*** SEQUENCING IN PROGRESS	
VERSION	AC005139.3	GI:4558581			
KEYWORDS	HTG; HTGS-PHASE1				
SOURCE	ORGANISM	malaria parasite P. falciparum,			
REFERENCE	AUTHORS	Plasmidium falciparum			
REFERENCE	AUTHORS	Ekaryota; Alveolata; Aplcomplexa; Haemosporida; Plasmodium,			
TITLE	JOURNAL	Hyman, R.W., Fung, E.L., Qin, F., Tamaki, T., Kurdi, O.B. and Davis, R.W.			
TITLE	JOURNAL	Plasmodium falciparum 3D7 chromosome 12			
TITLE	JOURNAL	Unpublished			
TITLE	JOURNAL	2 (bases 1 to 256172)			
TITLE	JOURNAL	Hyman, R.W., Qin, F., Fung, E.L., Conway, A.B. and Davis, R.W.			
TITLE	JOURNAL	Direct Submission			
TITLE	JOURNAL	Submitted (18-JUN-1998) Stanford DNA Sequencing and Technology			

```

COMMENT
On Apr 2, 1999, this sequence version replaced g1:4337170.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 126535: contig of 126535 bp in length
* 126536 126735: gap of unknown length
* 126736 158845: contig of 32110 bp in length
* 158846 159045: gap of unknown length
* 159046 211538: contig of 52433 bp in length
* 211539 248106: gap of unknown length
* 248107 248307: gap of unknown length
* 248307 256172: contig of 7866 bp in length.
*
* Location/Qualifiers
*   1..256172
*     /organism="Plasmodium falciparum"
*     /db_xref="taxon:5833"
*     /chromosome="12"
BASE COUNT      100700 a 26900 c 25826 g 101937 t 809 others
ORIGIN

```

	Query Match	14.7%	Score 52.8;	DB 60;	Length 256172;	
	Best Local Similarity	53.1%;	Pred. No. 4.1;			
	Matches 136;	Conservative 0;	Mismatches 117;	Indels 3;	Gaps 1;	
OY	1	aaaaaaaaaaagatgaaaglatcttattcctccctctcttttttgtgaataatttcaac	60			
Db	148285	AAAAATTAATAAACCAAMPAACCTTATTATTGTTGTTTCCTTATTATATAAATA	148344			
OY	61	tttttttgcc---caatgatatataaaaattggataataactatctggatactgc	117			
Db	148345	ATTTTTTGCAATTTAAAAAAACAATAAAAAATGTATATATTTGCATATATATTTT	148404			
OY	118	ttttaagtctgggtttggaagaaggcttcgaacttcgaaagtgacagtgatataag	177			
Db	148405	AATTTATTTTTTTTAAACAAAATATATATTTACAGCTTTCATGAGGAATAATAT	148464			
OY	178	tgggaagctaaggtctgagctcttcgacaattcgatcgaatcgtcatcataagtcg	237			
Db	148465	TTCGAGTAAATGGTTGTTATTTACCAAATTTGTAATCATATTTTAGATTAGC	148524			
OY	238	actattaacccttaa	253			
Db	148525	TTTATTTTAAATATPAA	148540			
RESULT 11						
LOCUS	AC005140					
DEFINITION	AC005140	310779 bp	DNA	HMG	12-AUG-2000	
ACCESSION	Plasmodium falciparum chromosome 12 clone 3D7, *** SEQUENCING IN					
VERSION	AC005140					
KEYWORDS	AC005140.8 GI:9797735					
SOURCE	HMG; HTGS_PHASE1.					
ORGANISM	malaria parasite P. falciparum.					
REFERENCE	Plasmodium falciparum					
AUTHORS	Eukaryotes; Alveolata; Apicomplexa; Haemosporida; Plasmodium.					
	1 (bases 1 to 310779)					
TITLE	Hyman,R.W., Fung,E.L., Qin,F., Rowley,D., Mao,J., Tamaki,T.,					
JOURNAL	Kurdi,O.B., Conway,A.B. and Davis,R.W.					
AUTHORS	Plasmodium falciparum 3D7 chromosome 12					
	Unpublished					
	2 (bases 1 to 310779)					
REFERENCE	Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.					
TITLE	Direct Submission					



JOURNAL Submitted (18-JUN-1998) Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

COMMENT On Aug 12, 2000 this sequence version replaced gi:8810455.  
 \* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

FEATURES  
 source  
 1. 187308: contig of 187308 bp in length  
 \* 187309 187508: gap of unknown length  
 \* 187509 257820: contig of 70312 bp in length  
 \* 257821 258020: gap of unknown length  
 \* 258021 307235: contig of 49235 bp in length  
 \* 307236 307455: gap of unknown length  
 \* 307456 310779: contig of 3324 bp in length.  
 Location/Qualifiers  
 1. 310779  
 /organism="Plasmodium falciparum"  
 /db\_xref="taxon:5833"  
 /chromosome="12"  
 /clone="PFYACB8-420"  
 /clone="3D7"

BASE COUNT 126134 a 29409 c 30056 g 124579 t 601 others

ORIGIN

Query Match 14.7%; Score 52.8; DB 60; Length 310779;  
 Best Local Similarity 53.1%; Pred. No. 4.1;  
 Matches 136; Conservative 0; Mismatches 117; Indels 3; Gaps 1;

OY 1 aaaaaaaaaaagatgaagatttttattctctcttttttttttgataatttaacaa 60  
 Db 7626 AAAATTTATATAACCAATTAACCTTATTTGTTCTCTTATTTATATAAATA 7685

OY 61 tttttttgc---caatgatataaaatttgataataatttttgatttgcgt 117  
 Db 7686 AATTTTGTGCAATTTAAAAAACATAAAAATGTAATATTTGATATATATTTT 7745

OY 118 ttttaagtcgggttggaaggggttcgaacttcgaagaggaagatgataagat 177  
 Db 7746 AATTTATTTTAAAAAATATATATATTCAGTTTCATGAGAAATATATATATTA 7805

OY 178 tggagagcaggttcagtccttggaacatttgatggatgcttgattatagtcgac 237  
 Db 7806 TTGGAGTTAATTTGTTGATATTTACAAATTTGTAATCATATTTTACATTAA 7865

OY 238 actataaaccttaa 253  
 Db 7866 TTTATTTATATATTA 7881

RESULT 12  
 AC014610 24091 bp DNA 16-NOV-1999  
 LOCUS AC014610  
 DEFINITION Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered pieces.  
 AC014610  
 AC014610.1 GI:6436725  
 VERSION HTG: HTGS PHASE2.  
 KEYWORDS HTG: HTGS PHASE2.  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 24091)  
 REFERENCE Adams, M. and Venter, J.C.  
 AUTHORS Direct Submission  
 TITLE Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA

COMMENT This sequence was identified as CDM:10214343 by the submitter.  
 For further information on this sequence e-mail to fly@celera.com.  
 \* NOTE: This is a 'working draft' sequence.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

FEATURES  
 source  
 1. 24091  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"

BASE COUNT 7749 a 4270 c 4283 g 7789 t

ORIGIN

Query Match 14.7%; Score 52.6; DB 63; Length 24091;  
 Best Local Similarity 46.8%; Pred. No. 5.5;  
 Matches 166; Conservative 0; Mismatches 189; Indels 0; Gaps 0;

OY 1 aaaaaaaaaaagatgaagatttttattctctcttttttttttgataatttaacaa 60  
 Db 18191 AAAAGTCAGAAAGTCATATTAATAAAAAAAACGCTGTATTTTATATATATAT 18250

OY 61 ttttttgcacaaatgatataaaatttgataataattatgataatttcgttt 120  
 Db 18251 ATATATATAGAAATATATATATTTTAAACAAATGATGCTTATATATACAGTTT 18310

OY 121 ttatgtcgggttggaaggggttcgaacttcgaaagtgagcagatgataagattg 180  
 Db 18311 GAAAAAATTTTAAAGTGTTTATTTATTTTGAATAATGATCTTAAATATTTGT 18370

OY 181 gagcaggttcgcttcggacatttgatcgttcgttgatcattatagtcgacact 240  
 Db 18371 AATTAATTTTTCACCTTTGTAATTAAGATCGATTTATATAAAGTCATTTTCG 18430

OY 241 attaaccttaaatgggttcctataagggccaattatatacgaattataacaaatgac 300  
 Db 18431 GCCAGACCAAAAGTCCGTTTACGAATCGAAAAATTTTACGATTTGAAAAAG 18490

OY 301 aactttacttcgttttttgatccgaagcaataacaaatgtaacaaacac 355  
 Db 18491 CATGATTTTCTATTTTCAATGAATTTGAAATTCCTTACGAATTCGATTTAC 18545

RESULT 13  
 AE002751 106993 bp DNA 06-OCT-2000  
 LOCUS AE002751  
 DEFINITION Drosophila melanogaster genomic scaffold 142000013386031, complete sequence.  
 AE002751  
 AE002751.2 GI:10729377  
 VERSION HTG: HTGS PHASE2.  
 KEYWORDS HTG: HTGS PHASE2.  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 106993)  
 REFERENCE Adams, M.D., Celisner, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galie, R.F., George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N., Sutton, G.C., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X., Brandon, R.C., Rogers, Y.H., Blazey, R.G., Champe, M., Pfeiffer, B.D., Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor Miklos, G.L., Abril, J.F., Adayant, A., An, H.J., Andrews, Pannkock, C., Baldwin, D., Ballew, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y., Benos, P.V., Bernier, B.P., Bhandari, D., Bolishakov, S., Borikova, D., Botchan, M.R., Bouck, J., Brockstein, P., Brothier, P., Burdick, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J.M., Crawley, S., Dahlke, C., Davenport, L.B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M., Dodson, K., Dou, P.L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Fiertera, S.,

Flieschmann, W., Foster, C., Gabriellian, A.E., Garg, N.S., Gelbalt, W.M., Glasser, K., Glodok, A., Gong, F., Gotrell, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Helman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpén, G.H., Ke, Z., Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasco, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Matel, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Mishina, N.V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusser, D.R., Pacle, J.M., Palazolo, M., Peltman, G.S., Pan, S., Pollard, J., Pui, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Svirska, R., Tector, C., Turner, R., Venter, L., Wang, A.H., Wang, X., Wang, Z.Y., Wasserman, D.A., Weinstein, G.M., Weissbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yen, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.  
The genome sequence of *Drosophila melanogaster*  
Science 287 (5461), 2185-2195 (2000)

2 (bases 1 to 106993)  
Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J.  
Direct Submission  
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA  
On Oct 9, 2000 this sequence version replaced gi:7289853.  
Location/Qualifiers  
1. 106993  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/chromosome="2R"  
/map="41C-41D"  
/map="41F1-41F1"  
join(<57498, .57614, 57686, .>57928)  
/gene="CG17482"  
/product="CT38651"  
/db\_xref="FLYBASE:FBan0017482"  
/db\_xref="FLYBASE:FBgn0032994"  
<57498, .>57928  
/gene="CG17482"  
/db\_xref="FLYBASE:FBan0017482"  
/db\_xref="FLYBASE:FBgn0032994"  
join(57498, .57614, 57686, .57928)  
/gene="CG17482"  
/note="CG17482 gene product"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0017482"  
/db\_xref="FLYBASE:FBgn0032994"  
/protein\_id="AAFA5460.1"  
/db\_xref="GI:7289853"  
/translation="MDAGSIIIVCYMSRTGSNDMVTGKTNETSEQYPELCTARDSND  
ERNPDYEMEVMTFGASCSPSSAHNVKTLTLKFMPSDSRAVFTIDYHYVDYVSF  
VTERSSPYEYELGPG"  
join(68140, .68292, 72127, .72917, 79504, .80049, 81286, .82015,  
87046, .87370)  
/gene="p120ctn"  
/product="CT38657"  
/db\_xref="FLYBASE:FBan0017484"  
/db\_xref="FLYBASE:FBgn0015587"  
<68140, .>87370  
/gene="p120ctn"  
/note="CG17484"  
/db\_xref="FLYBASE:FBan0017484"  
/db\_xref="FLYBASE:FBgn0015587"  
join(72160, .72917, 79504, .80049, 81286, .82015, 87046, .87351)  
/gene="p120ctn gene product"  
/note="p120ctn gene product"  
/codon\_start=1

/db\_xref="FLYBASE:FBan0017484"  
/db\_xref="FLYBASE:FBgn0015587"  
/protein\_id="AAFA5461.1"  
/db\_xref="GI:7289855"  
/translation="MQSORGFQWYDDPSAAGYSSTPLTYTKPTGCAITKRVAPHC  
NKITNSTSDSPIDENHPLATVRYVQAGYEDTSEGVAGLTCGIDSETYTAAVP  
YCYSDANVYGIQNTTSIKPFSPRPNDNGEAVADSCRTFKSAEKLPOFM  
SSINTVQRLKEKDYIEGSENDICPMRNDPLISYIFLSNPSSAIFAMAAYLQ  
HICMDPNDKORTSLGIPLYLVLISYDSPEIKNCGALRNLSYRONDENRGIK  
NAGCIATLHLCRSQETPEYKELVYGLMNSCEDIKRSITIDELAAVCSYIKPK  
GMDVACGECFSPVFNMAAGVILNYSAGEHAACCLRNCEHLYECILYVYSIEKN  
NIGKIVNCECILRNLSYRQEVDDNYDHPITPERYIPSCNPEETETAAAGLQ  
NLSACYQPSIDIRATVKEKGLIIVELMEYDRAVCAVATLRNLAIIDQNRKL  
GKYMRDLVQKPSGVNOHQDQNSDDITAVLAINVINKNKEODPFVSKHFTANTPS  
MNIKREKETSCLKPAQVLYTMMQHNELRDYKKNKKEODPFVSKHFTANTPS  
SPNVNVLNRPMAOSQRTREDTIOGSTYLSANDSGAVVNSAMISEKRA  
EMQKVNDVIAIRLYQYEDGILKIRKGLKIAKVIKHAHYDEVKDGDEGPKTLT  
VSEPKAFEGINTQSGLPVIGKKEPQIGRETRSGRL"  
complement(join(<75533, .75970, 76088, .76363))  
/gene="CG17479"  
/product="CT38649"  
/db\_xref="FLYBASE:FBan0017479"  
/db\_xref="FLYBASE:FBgn0032996"  
complement(<75533, .>76363)  
/gene="CG17479"  
/db\_xref="FLYBASE:FBan0017479"  
/db\_xref="FLYBASE:FBgn0032996"  
complement(join(75692, .75970, 76088, .76363))  
/gene="CG17479"  
/note="CG17479 gene product"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0017479"  
/db\_xref="FLYBASE:FBgn0032996"  
/protein\_id="AAFA5464.1"  
/db\_xref="GI:7289858"  
/db\_xref="GI:7289858"  
/translation="MASIAVRANTENSKYSVEVRAGEDTGAEGYRYTPGIRLENG  
EGSSIRSFAGIKTVPSKGLSQSPAALSKRIQWIKRDEEHLEFTAVRDL  
KTEVALSETRYRVNPGCKWVSDQKGTISTAVIGARTISDDAAIVIAGLVFDELY  
LKKDARRSSLDNOMAMWNSKGR"  
91799, .>93503  
/gene="CG17486"  
/product="CT38665"  
/db\_xref="FLYBASE:FBan0017486"  
/db\_xref="FLYBASE:FBgn0032997"  
<91799, .>93503  
/gene="CG17486"  
/db\_xref="FLYBASE:FBan0017486"  
/db\_xref="FLYBASE:FBgn0032997"  
91809, .93503  
/gene="CG17486"  
/note="CG17486 gene product"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0017486"  
/db\_xref="FLYBASE:FBgn0032997"  
/protein\_id="AAFA5462.1"  
/db\_xref="GI:7289856"  
/translation="MCGICFCSVNNVPLNFSNLSALVKVILKNRGPDDVYDYCF  
GKILAFGEVLMQGESYQKQPVYDDPIELFNGDIYNTSKPEYMSOTMTAELAEGR  
CKEONIKILKREGECHLITYKRBIDILFSDDALCRNSULLERICNGQLSTSHF  
EDRKISLEDPPLGLFRVKNLDSVLYPMQJLNSSTOLRLNLDLAIGKTAVESW  
SPDMWLNCPNPFYSNFEYQIKNNVLYNLIDQSEIKYTLISLHLTLSDSINKNR  
NMAFCRLQKMLNISPGRHAKCLIFSGGLCTTIALANQVFNPEPILINLVAF  
EYKGVONISEKLFDPDRKRLSVNVEIKQCEPCWNLKAVVTYRLQOHLTSRK  
HLIYPLDYVDESIGCAFWASCHTSTAVAILIGSGADELFGGYTFRNYSRQCN  
DIEROLAVONELEERDMORISARNLARDRIATGCTARSPTIEENRYKTRISLEYO  
KCCGFPEGVGDLLRLRYGQIGLDRVFLKRLAIQFSGRIANKQNGSHSDNL"  
complement(join(93644, .93777, 94377, .94539, 94544, .>95261))  
/gene="CG17478"  
/product="CT38647"  
/db\_xref="FLYBASE:FBan0017478"  
/db\_xref="FLYBASE:FBgn0032998"  
complement(<93644, .>95261)  
/gene="CG17478"











```

FT      Intron      4067..4164
FT      /tag= 1
FT      exon      4165..4479
FT      /tag= 1
FT      Intron      4480..4631
FT      /tag= k
FT      exon      4632..4724
FT      /tag= 1
FT      Intron      4725..4815
FT      /tag= m
FT      exon      4816..4894
FT      /tag= n
FT      Intron      4895..5000
FT      /tag= o
FT      exon      5001..5110
FT      /tag= p
FT      Intron      5111..5864
FT      /tag= q
FT      exon      5865..6110
FT      /tag= r
FT      3'UTR      6011..6468
FT      /tag= s
XX      WO200047715-A2.
XX      17-AUG-2000.
XX      11-FEB-2000; 2000WO-US03820.
XX      11-FEB-1999; 99US-0119657.
XX      11-FEB-1999; 99US-0119658.
XX      (ARIZ-) ARIZONA BOARD OF REGENTS.
XX      PI      Azpiroz R, Choe S, Feldmann KA;
XX      DR      WPI: 2000-549142/50.
XX      DR      P-PSDB; AAB07921.
XX      PT      New isolated dwf4 polynucleotide useful for altering the phenotype of
XX      PT      plants, for diagnostic assays and in the production of antibodies -
XX      PS      Claim 3; Fig 10A-G; 113pp; English.
XX      CC      The present sequence encodes a DWF4 polypeptide. The polypeptide is a
XX      CC      cytochrome P450 enzyme that mediates multiple steps in synthesis of
XX      CC      brassinosteroids. Specifically, it mediates multiple
XX      CC      22alpha-hydroxylation steps in brassinosteroid biosynthesis. The DWF4
XX      CC      polynucleotide is used for altering the phenotype of a plant. DWF4
XX      CC      plants display a dramatic reduction in the length of different organs,
XX      CC      and this size reduction is attributable to a defect in cell elongation.
XX      CC      The DWF4 polynucleotides and polypeptides can be used in diagnostic
XX      CC      assays and to generate antibodies, which can be used to produce
XX      CC      immunogenic compositions.
XX      SQ      Sequence 6888 BP; 2294 A; 1010 C; 1193 G; 2391 T; 0 other;

```

```

Query Match      100.0%; Score 358; DB 21; Length 6888;
Best Local Similarity 100.0%; Pred. No. 8.7e-56;
Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy      1 aaaaaaaaaagatgaagatatttatctctctcttttttttttgataatttaaca 60
Db      6111 aaaaaaaaaagatgaagatatttatctctctcttttttttttgataatttaaca 6170
Oy      61 tttttttgcccaatgataataaaatttgataataatataatgataatcgcttt 120
Db      6171 tttttttgcccaatgataataaaatttgataataatataatgataatcgcttt 6230
Oy      121 ttacttcgggttgagaaaagggttcgacttcgaaagtgcagcatgtatatagatgg 180
Db      6231 ttacttcgggttgagaaaagggttcgacttcgaaagtgcagcatgtatatagatgg 6290

```

```

Oy      181 gagctaggctgagctcttgagacatttgatggatggttgatattatagtgacac 240
Db      6291 gagctaggctgagctcttgagacatttgatggatggttgatattatagtgacac 6350
Oy      241 attaaccttaaatgagctcttctataaaggcccaatataatataatgataataaagtgac 300
Db      6351 attaaccttaaatgagctcttctataaaggcccaatataatataatgataataaagtgac 6410
Oy      301 aactttacttcggttttttgatccggaagcaataacaattgtcaatataccaacaag 358
Db      6411 aactttacttcggttttttgatccggaagcaataacaattgtcaatataccaacaag 6468

```

## RESULT 2

```

ID      AAF58252/C
ID      AAF58252 standard; DNA; 936 BP.
XX      AAF58252;
XX      24-APR-2001 (first entry)
XX      Oligonucleotide D1835.
XX      Electron-Transfer group; ETM; mismatch; genotyping;
XX      gene expression; ss.
XX      Synthetic.
XX      WO200107665-A2.
XX      01-FEB-2001.
XX      26-JUL-2000; 2000WO-US20476.
XX      26-JUL-1999; 99US-0145695.
XX      17-MAR-2000; 2000US-0190259.
XX      (CLIN-) CLINICAL MICRO SENSORS INC.
XX      PI      Umek RM;
XX      DR      WPI: 2001-159728/16.
XX      PT      Nucleic acids containing electron-transfer group, useful as labels in
XX      PT      hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX      PT      a single surface -
XX      PS      Example 6; Page 127; 159pp; English.
XX      CC      The present invention relates to a composition comprising two nucleic
XX      CC      acids each containing an electron-transfer group (ETM) having
XX      CC      different redox potentials. The invention is used for electronic
XX      CC      detection of nucleic acids, especially of substitutions (mismatches)
XX      CC      and single-nucleotide polymorphisms, e.g. for genotyping,
XX      CC      monitoring gene expression.
XX      SQ      Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

```

```

Query Match      27.5%; Score 98.6; DB 22; Length 936;
Best Local Similarity 1.4%; Pred. No. 1e-09;
Matches 5; Conservative 254; Mismatches 98; Indels 0; Gaps 0;

```

```

Oy      1 aaaaaaaaaagatgaagatatttatctctctcttttttttttgataatttaaca 60
Db      6011 aaaaaaaaaagatgaagatatttatctctctcttttttttttgataatttaaca 602
Oy      61 tttttttgcccaatgataataaaatttgataataatataatgataatcgcttt 120
Db      6011 tttttttgcccaatgataataaaatttgataataatataatgataatcgcttt 542
Oy      121 ttacttcgggttgagaaaagggttcgacttcgaaagtgcagcatgtatatagatgg 180

```



[illegible]

RESULT	3
AAF58254/C	
ID	AAF58254 standard; DNA; 936 BP.

DT	24-APR-2001	(first entry)
XX		
DE	Oligonucleotide D1875.	
XX		

KW Electron-transfer group; ETM; mismatch; genotyping;  
KW gene expression; ss.

Synthetic.

PN WO200107665-A2.

PD 01-FEB-2001

26-JUL-2000; 2000WO-US20476.

PR 26-JUL-1999; 99US-0145695.

[illegible]

2000

PI Umek RM;  
VY

DR WPI; 2001-159728/16.

XX      a single surface

PT      nucleic acids containing electron-transfer group, useful as labels in hybridization assays, e.g., for genotyping, allowing repeat analyses on

PS      Example 6; Page 127; 159pp; English.

CC The present invention relates to a composition comprising two nucleic  
CC acids each containing an electron-transfer group (ETM) having  
CC different redox potentials. The invention is used for electronic  
CC detection of nucleic acids, especially of substitutions (mismatches)  
CC and single-nucleotide polymorphisms, e.g. for genotyping,  
CC monitoring gene expression.

SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match	27.5%	Score 98.6	DB 22	Length 936
Best Local Similarity	1.4%	pred. No. 1e-09		
Matches	5	Conservative 254	Mismatches 98	Indels 0
				Gaps 0

Oy	1	aaaaaaaaagaagaaagtatttttcctccttlttttttgtaatttaaaca	60
Db	661	mmmw	602
Oy	61	ttttttggcccaatgatataaaaaattggaataaatatatatgtgcatttgtt	120
Db	601	mwmwmwmmmcmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmw	542

Oy 121 ttgttcggttttgagaaggtttgcacttcgaagcagcagatataagattg 180  
 Db 541 ww 482  
 Oy 181 gacgaagttgagctcttgacatttgatggaattgttgatataatgtgcgaact 240  
 Db 481 ww 422  
 Oy 241 attaaaccttaatggccttctcttaagcccaattatataagattatacaaatgtac 300  
 Db 421 ww 362  
 Oy 301 aaccttacttcgttttgatcgcgaagcaataacacaatttgcataaccacaacaa 357  
 Db 361 ww 305

RESULT	4
AAFS8257/c	
ID	AAFS8257 standard; DNA; 936 BP.
vv	

AC AAF58257;  
yy

DT 24-APR-2001 (first entry)  
 YY

Oligonucleotide D1954.

**KM** Electron-transfer group; ETM; mismatch; genotyping, gene expression; ss

XX  
XX  
Synthetic

XX WO200107665-A2  
PN

XX PD 01-FEB-2001

26-JUL-2000 : 2000WO-11S20476

XX 26-JUL-1999: 09115-0145695  
PRPR 1 / -MAR-2000; 200005-0190259.  
XXPA (CLIN-) CLINICAL MICRO SENSORS INC.  
XX  
XX

P1	Umek RM;
XX	

WFL; 2001-159/28/1b-  
XX  
XX

PT Nucleic acids containing electron-transfer group, useful as labels in  
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
PT a single surface -  
XX  
XX  
PS Example 6; Page 127; 159pp; English.  
XY

CC The present invention relates to a composition comprising two nucleic  
CC acids each containing an electron-transfer group (ETM) having  
CC different redox potentials. The invention is used for electrocyclic  
CC detection of nucleic acids, especially of substitutions (mismatches)  
CC and single-nucleotide polymorphisms, e.g. for genotyping,  
CC monitoring gene expression.

Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match	27.5%;	Score	98.6;	DB	22;	Length	936;
Best Local	Similarity	1.48;	Pred.	No. 1e-09;			
Matches	5;	Conservative	254;	Mismatches	98;	Indels	0;
						Gaps	0;

```

0y 1 aaaaaaaaaagatgaatgttttttaatcctctcttttttttggtaatttaaaaca 60
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
661 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 602
0y 61 tttttttgccaatgatataaaaaatttgataataatatatgatgatcgtttt 120

```

[illegible]

RESULT 5  
AAFS8259/C  
ID AAF58259 standard; DNA: 936 BP.  
XX  
XX AAF58259;  
AC  
XX  
DT 24-APR-2001 (first entry)  
XX  
DE Oligonucleotide D2004.  
XX  
XX Electron-transfer group; ETM; mismatch; genotyping;  
KM gene expression; ss.  
XX  
XX Synthetic.  
OS  
PN WO200107665-A2.  
XX  
XX 01-FEB-2001.  
PD  
XX  
XX 26-JUL-2000; 2000WO-US20476.  
PF  
XX 26-JUL-1999; 99US-0145695.  
PR 17-MAR-2000; 2000US-0190259.  
XX  
XX  
PA (CLIN-) CLINICAL MICRO SENSORS INC.  
XX  
XX Umek RM;  
PI  
DR WPI: 2001-159728/16.  
XX  
XX  
PT Nucleic acids containing electron-transfer group, useful as labels in  
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
PT a single surface -  
PT  
XX  
PS Example 6; Page 128; 159pp; English.  
XX  
XX  
CC The present invention relates to a composition comprising two nucleic  
CC acids each containing an electron-transfer group (ETM) having  
CC different redox potentials. The invention is used for electronic  
CC detection of nucleic acids, especially of substitutions (mismatches)  
CC and single-nucleotide polymorphisms, e.g. for genotyping.  
XX monitoring gene expression.  
XX  
XX Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

[illegible][illegible]

RESULT 6  
 AAF58262/C  
 ID AAF58262 standard; DNA; 936 BP.  
 XX  
 AC AAF58262;  
 XX  
 DT 24-APR-2001 (first entry)  
 XX  
 DE Oligonucleotide D2007.  
 XX  
 KW Electron-transfer group; ETM; mismatch; genotyping;  
 KW gene expression; ss.  
 XX  
 OS Synthetic.  
 PN WO200107665-A2.  
 XX  
 PD 01-FEB-2001.  
 XX  
 PF 26-JUL-2000; 2000WO-US20476.  
 XX  
 PR 26-JUL-1999; 99US-0145695.  
 PR 17-MAR-2000; 2000US-0190259.  
 XX  
 PA (CLIN-) CLINICAL MICRO SENSORS INC.  
 XX  
 P1 Umek RM;  
 XX  
 DR WPI; 2001-159728/16.  
 XX  
 PT Nucleic acids containing electron-transfer group, useful as labels in  
 PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
 PT a single surface -  
 XX  
 PS Example 6; Page 128; 159pp; English.  
 XX  
 CC The present invention relates to a composition comprising two nucleic  
 CC acids each containing an electron-transfer group (ETM) having  
 CC different redox potentials. The invention is used for electronic  
 CC detection of nucleic acids, especially of substitutions (mismatches)  
 CC and single-nucleotide polymorphisms, e.g. for genotyping.  
 CC monitoring gene expression.  
 XX  
 SN Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;  
 XX

Query Match	27.5%	Score 98.6;	DB 22;	Length 936;
Best Local Similarity	1.48;	Pred. No. 1e-09;		
Matches	5;	Conservative 254;	Mismatches 98;	Indels 0;
				Gaps 0;
0Y	1	aaaaaaaaaagatgaagatattttattctctctcttttttttttgaataatttaacaa	60	

```
Db 661 WWWWWW.....:
Oy 61 ttttttggcccaatgatatataaaattggataataatttgcgttt 120
Db 601 WWWWWW.....:
Oy 121 ttatgtcggtttggaaaagggttcgacttcgaaggtagacgtatataatgg 180
Db 541 WWWWWW.....:
Oy 181 gaggcaggttgagcttcggacatttgatggatgttgattatagtcgacact 240
Db 481 WWWWWW.....:
Oy 241 attaaccttaaatgagcttcctataaggccaattatacgtataaacaagtgc 300
Db 421 WWWWWW.....:
Oy 301 aactttacttcgttttggatccgagcaataacaattgtcaataacacacaa 357
Db 361 WWWWWW.....:

RESULT 7
AAF58255/c
ID AAF58255 standard; DNA; 938 BP.
AC AAF58255;
XX
XX
XX 24-APR-2001 (first entry)
DT
DE Oligonucleotide D1876.
XX
XX Electron-transfer group; ETM; mismatch; genotyping;
KM gene expression; ss.
XX
XX Synthetic.
OS
XX WO200107665-A2.
XX
XX 01-FEB-2001.
PD
XX
XX 26-JUL-2000; 2000WO-US20476.
PE
XX
XX 26-JUL-1999; 99US-0145695.
PR
XX 17-MAR-2000; 2000US-0190259.
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
XX Umek RM;
PI
XX WPI; 2001-159728/16.
DR
XX
XX Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
XX Example 6; Page 127; 159pp; English.
XX
XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
XX Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;
SQ
```

Query Match 27.5%; Score 98.6; DB 22; Length 938;  
Best Local Similarity 1.4%; Pred. No. 1e-09;  
Matches 5; Conservative 254; Mismatches 98; Indels 0; Gaps 0;

```
Oy 1 aaaaaaaaaatgaatattttatctctcttttttttgataatttaaca 60
Db 661 WWWWWW.....:
Oy 61 ttttttggcccaatgatatataaaattggataataatttgcgttt 120
Db 601 WWWWWW.....:
Oy 121 ttatgtcggtttggaaaagggttcgacttcgaaggtagacgtatataatgg 180
Db 541 WWWWWW.....:
Oy 181 gaggcaggttgagcttcggacatttgatggatgttgattatagtcgacact 240
Db 481 WWWWWW.....:
Oy 241 attaaccttaaatgagcttcctataaggccaattatacgtataaacaagtgc 300
Db 421 WWWWWW.....:
Oy 301 aactttacttcgttttggatccgagcaataacaattgtcaataacacacaa 357
Db 361 WWWWWW.....:

RESULT 8
AAF58252
ID AAF58252 standard; DNA; 936 BP.
AC AAF58252;
XX
XX
XX 24-APR-2001 (first entry)
DT
DE Oligonucleotide D1835.
XX
XX Electron-transfer group; ETM; mismatch; genotyping;
KM gene expression; ss.
XX
XX Synthetic.
OS
XX WO200107665-A2.
XX
XX 01-FEB-2001.
PD
XX
XX 26-JUL-2000; 2000WO-US20476.
PE
XX
XX 26-JUL-1999; 99US-0145695.
PR
XX 17-MAR-2000; 2000US-0190259.
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
XX Umek RM;
PI
XX WPI; 2001-159728/16.
DR
XX
XX Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
XX Example 6; Page 127; 159pp; English.
XX
XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
XX Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;
SQ
```

Query Match 27.1%; Score 97; DB 22; Length 936;

Best Local Similarity 1.1%; Pred. No. 1.9e-09;  
Matches 4; Conservative 254; Mismatches 99; Indels 0; Gaps 0;

```
Oy 1 aaaaaaaaaagatgaagatttttattctctcttttttttgataatttaacaa 60
Db 324 www.....
Oy 61 ttttttgcacatgatatataaaatttgataataatatttgatattcgtttt 120
Db 364 www.....
Oy 121 ttacttcgggtttgagaaagggttcgacttcgaaagtgaagatgatatagattg 180
Db 444 www.....
Oy 181 gagcaggttgagctcttgagacatttgatcgatgctgtgatataagtgacact 240
Db 504 www.....
Oy 241 attaaccttaaatgagcttctataagcccaattatatacgattatacaaaagtac 300
Db 564 www.....
Oy 301 aactttacttcgtttttgatccgaagcaataacaaattgtcaaataccacaa 357
Db 624 www.....
```

## RESULT 9

AAFS8254  
ID AAF58254 standard; DNA; 936 BP.

AC AAF58254;

DT 24-APR-2001 (first entry)

DE Oligonucleotide D1875.

DE Electron-transfer group; ETM; mismatch; genotyping;

KW gene expression; ss.

OS Synthetic.

PN MO200107665-A2.

PD 01-FEB-2001.

PD 26-JUL-2000; 2000MO-US20476.

PF 26-JUL-1999; 99US-0145695.

PR 17-MAR-2000; 2000US-0190259.

PR (CLIN-) CLINICAL MICRO SENSORS INC.

PA Umek RM;

PI WPI: 2001-159728/16.

PT Nucleic acids containing electron-transfer group, useful as labels in

PT hybridization assays, e.g. for genotyping, allowing repeat analyses on

PT a single surface

PS Example 6; Page 127; 159pp; English.

PS The present invention relates to a composition comprising two nucleic

PS acids each containing an electron-transfer group (ETM) having

PS different redox potentials. The invention is used for electronic

PS detection of nucleic acids, especially of substitutions (mismatches)

PS and single-nucleotide polymorphisms, e.g. for genotyping,

PS monitoring gene expression.

Query Match 27.1%; Score 97; DB 22; Length 936;  
Best Local Similarity 1.1%; Pred. No. 1.9e-09;  
Matches 4; Conservative 254; Mismatches 99; Indels 0; Gaps 0;

```
Oy 1 aaaaaaaaaagatgaagatttttattctctcttttttttgataatttaacaa 60
Db 324 www.....
Oy 61 ttttttgcacatgatatataaaatttgataataatatttgatattcgtttt 120
Db 364 www.....
Oy 121 ttacttcgggtttgagaaagggttcgacttcgaaagtgaagatgatatagattg 180
Db 444 www.....
Oy 181 gagcaggttgagctcttgagacatttgatcgatgctgtgatataagtgacact 240
Db 504 www.....
Oy 241 attaaccttaaatgagcttctataagcccaattatatacgattatacaaaagtac 300
Db 564 www.....
Oy 301 aactttacttcgtttttgatccgaagcaataacaaattgtcaaataccacaa 357
Db 624 www.....
```

## RESULT 10

AAFS8257  
ID AAF58257 standard; DNA; 936 BP.

AC AAF58257;

DT 24-APR-2001 (first entry)

DE Oligonucleotide D1954.

DE Electron-transfer group; ETM; mismatch; genotyping;

KW gene expression; ss.

OS Synthetic.

PN MO200107665-A2.

PD 01-FEB-2001.

PD 26-JUL-2000; 2000MO-US20476.

PF 26-JUL-1999; 99US-0145695.

PR 17-MAR-2000; 2000US-0190259.

PR (CLIN-) CLINICAL MICRO SENSORS INC.

PA Umek RM;

PI WPI: 2001-159728/16.

PT Nucleic acids containing electron-transfer group, useful as labels in

PT hybridization assays, e.g. for genotyping, allowing repeat analyses on

PT a single surface

PS Example 6; Page 127; 159pp; English.

PS The present invention relates to a composition comprising two nucleic

PS acids each containing an electron-transfer group (ETM) having

PS different redox potentials. The invention is used for electronic

PS detection of nucleic acids, especially of substitutions (mismatches)

PS and single-nucleotide polymorphisms, e.g. for genotyping,

PS monitoring gene expression.

CC and single-nucleotide polymorphisms, e.g. for genotyping,

CC different redox potentials. The invention is used for electronic

CC detection of nucleic acids, especially of substitutions (mismatches)  
CC and single-nucleotide polymorphisms, e.g. for genotyping,  
CC monitoring gene expression.

SQ Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;

Query Match	27.1%	Score 97	DB 22	Length 936
Best local	Similarity 1.1%	Pred. No. 1.9e-09		
Matches	4	Conservative	254	Mismatches 99; Indels 0; Gaps 0

QY 1 aaaaaaaaaagaagaaagtaatttccttcctcttcttttttttgaataattttaaaca 60  
324 www 384  
QY 61 tttttttgccaatgatataataaatttgcatacaataatatttgcataatcgtttt 120  
384 wwwwwwwwwggcttaawww 444  
QY 121 ttaattcggatttagaaaaagggttcgacttcgaaagtgcagatgatataatggtg 180  
444 www 504  
QY 181 gagctaggttgaagctcttgcgcatttgatggagtgtgtgatattagtgcgcacat 240  
504 www 564  
QY 241 attaaccttaaatgggcttctctaagggcccaattatcagattaaacaagtac 300  
564 www 624  
QY 301 aactttactcgtttttgatcgaagaacaataacaatttgcataatccaaaccaa 357  
624 www 680

## RESULT 13

ID	AAF58255 standard; DNA; 938 BP
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

AC AAF58255;

DT 24-APR-2001 (first entry)

DE Oligonucleotide D1876.

KW Electron-transfer group; ETM; mismatch; genotyping,  
KW gene expression; ss.

OS Synthetic.

PN WO200107665-A2

PD 01-FEB-2001

PF 26-JUL-2000; 2000WO-US20476

PR 26-JUL-1999; 99US-0145695.

[illegible]

PA (CLIN-) CLINICAL MICRO SENSORS INC

PI Umek RM;

DR WPI; 2001-159728/16.

PT Nucleic acids containing electron-transfer group, useful as labels in hybridization assays, e.g. for genotyping, allowing repeat analyses on a single surface -

CC The present invention relates to a composition comprising two nucleic

CC acids each containing an electron-transfer group (ETM) having  
CC different redox potentials. The invention is used for electronic  
CC detection of nucleic acids, especially of substitutions (mismatches)  
CC and single-nucleotide polymorphisms, e.g., for genotyping,  
CC monitoring gene expression.

SQ Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other,

Query Match	27.1%	Score 97:	DB 22:	length 938:
Best local Similarity	1.1%	Pred No. 1.9e-09:		
Matches	4:	Conservative	99:	Indels 0:
		Mismatches	99:	Gaps 0

[illegible]

RESULT 14

ID AAF58238 standard; DNA; 244 BP.

AC AAF58238;

DT 24-APR-2001 (first entry)

DE Oligonucleotide D1250:D1102.

KW Electron-transfer group; ETM; mismatch; genotyping  
KW gene expression; ss.

OS Synthetic.

PN W0200107665-A2

PD 01-FEB-2001

PF 26-JUL-2000; 2000WO-US20476.

PR 26-JUL-1999; 99US-0145695.

PR 17-MAR-2000; 2000US-0190259.

PA (CLIN-) CLINICAL MICRO SENSORS INC.

PI Umek RM;

DR WPI; 2001-159728/16

PT Nucleic acids containing electron-transfer group, useful as labels in hybridization assays, e.g. for genotyping, allowing repeat analyses on a single surface -

PS Example 4; Page 120; 159pp; English.

XX The present invention relates to a composition comprising two nucleic  
CC acids each containing an electron-transfer group (ETM) having  
CC different redox potentials. The invention is used for electronic  
CC detection of nucleic acids, especially of substitutions (mismatches)  
CC and single-nucleotide polymorphisms, e.g. for genotyping,  
CC monitoring gene expression.  
XX  
SQ Sequence 244 BP; 19 A; 9 C; 12 G; 10 T; 194 other;

## Query Match

16.5%; Score 59.2; DB 22; Length 244;

Best Local Similarity 0.6%; Pred.No. 0.01;

Matches 1; Conservative 137; Mismatches 40; Indels 0; Gaps 0;

OY 1 aaaaaaaaaaagatgaagatatttattctctctcttttttttgataatttaataca 60  
Db 196 www 137  
OY 61 ttttttgcacatgatatataaaatttgataataatttgatatattcgttt 120  
Db 136 www 77  
OY 121 ttagtcgggttgagaaaagggttcgacttcgaaagtgacgacgatatagatt 178  
Db 76 www 19

## RESULT 15

AAFS8238  
ID AAF58238 standard; DNA; 244 BP.

XX AAF58238;

XX 24-APR-2001 (first entry)

XX Oligonucleotide D1250:D1102.

XX Electron-transfer group: ETM; mismatch; genotyping;

XX gene expression; ss.

XX Synthetic.

XX WO200107665-A2.

XX 01-FEB-2001.

XX 26-JUL-2000; 2000WO-US20476.

XX 26-JUL-1999; 98US-0145695.

XX 17-MAR-2000; 2000US-0190259.

XX (CLIN-) CLINICAL MICRO SENSORS INC.

XX Umek RM;

XX WPI; 2001-159728/16.

XX Nucleic acids containing electron-transfer group, useful as labels in  
XX hybridization assays, e.g. for genotyping, allowing repeat analyses on  
XX a single surface

XX Example 4; Page 120; 159pp; English.

XX The present invention relates to a composition comprising two nucleic  
XX acids each containing an electron-transfer group (ETM) having  
XX different redox potentials. The invention is used for electronic  
XX detection of nucleic acids, especially of substitutions (mismatches)  
XX and single-nucleotide polymorphisms, e.g. for genotyping,  
XX monitoring gene expression.

XX Sequence 244 BP; 19 A; 9 C; 12 G; 10 T; 194 other;

Query Match 16.2%; Score 58; DB 22; Length 244;

Best Local Similarity 4.0%; Pred.No. 0.017; Mismatches 42; Indels 0; Gaps 0;

Matches 7; Conservative 127; Mismatches 42; Indels 0; Gaps 0;

OY 1 aaaaaaaaaaagatgaagatatttattctctctcttttttttgataatttaataca 60  
Db 37 www 96  
OY 61 ttttttgcacatgatatataaaatttgataataatttgatatattcgttt 120  
Db 97 www 156  
OY 121 ttagtcgggttgagaaaagggttcgacttcgaaagtgacgacgatatagatt 176  
Db 157 www 212

Search completed: October 17, 2001, 19:19:16  
Job time: 15956 sec

Thu Oct 18 08:39:11 2001

us-09-502-426-1\_copy\_6111\_6468.rng

---

Page 10

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2001, 19:14:07 ; Search time 156 Seconds  
(without alignments)  
434.445 Million cell updates/sec

Title: US-09-502-426-1\_COPY\_6111\_6468  
Perfect score: 358  
Sequence: 1 aaaaaaaaaaagatgaagt.....tgtcaataaccacacacaaag 358

Scoring table: IDENTITY\_NIC  
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_NA:\*  
1: /cgn2\_6/prodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/prodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/prodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/prodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/prodata/2/ina/PCITUS\_COMB.seq:\*  
6: /cgn2\_6/prodata/2/ina/Backfilest1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	42.8	12.0	1493	1	US-08-340-820-24
C 2	42.8	12.0	1493	1	US-08-593-535-24
C 3	41.8	11.7	998	4	US-09-122-400B-5
C 4	41	11.5	2230	4	US-08-378-313-24
C 5	40.8	11.4	6243	2	US-09-056-075-1
C 6	40.8	11.4	19557	5	PCIT-0592-06300-1
C 7	40.6	11.3	1733	3	US-09-073-569-1
C 8	40.6	11.3	3095	6	5231168-1
C 9	40.4	11.3	10607	1	US-08-078-090-3
C 10	40	11.2	2755	3	US-08-749-522-2
C 11	39.4	11.0	289	1	US-08-341-568-3
C 12	39.4	11.0	289	2	US-08-911-020-3
C 13	39.4	11.0	19124	2	US-08-973-273-4
C 14	39	10.9	9048	3	US-08-487-826B-13
C 15	38.8	10.8	863	4	US-08-998-416-547
C 16	38.8	10.8	835	4	US-08-998-416-498
C 17	38.8	10.8	3138	1	US-07-867-106-4
C 18	38.2	10.7	2836	3	US-08-747-221B-24
C 19	38.2	10.7	2836	3	US-08-747-221B-26
C 20	38.2	10.7	5852	1	US-07-867-106-2
C 21	38.2	10.7	25604	3	US-08-998-416-701
C 22	38	10.6	701	4	US-08-693-457-3
C 23	38	10.6	2555	2	US-09-265-731-3
C 24	38	10.6	2555	4	US-08-393-625-16
C 25	38	10.6	8700	2	US-08-466-961A-16
C 26	38	10.6	8700	2	US-08-645-193B-18
C 27	38	10.6	8700	2	US-08-645-193B-18

C 28	37.8	10.6	1582	3	US-08-545-196B-10	Sequence 10, Appl
C 29	37.8	10.6	1582	3	US-08-545-196B-12	Sequence 12, Appl
C 30	37.6	10.5	2422	1	US-07-867-106-5	Sequence 5, Appl
C 31	37.6	10.5	3138	1	US-07-867-106-4	Sequence 4, Appl
C 32	37.6	10.5	5852	1	US-07-867-106-2	Sequence 2, Appl
C 33	37.6	10.5	6078	4	US-09-173-914-1	Sequence 1, Appl
C 34	37.4	10.4	2445	6	5215909-9	Patent No. 5215909
C 35	37.2	10.4	144	1	US-08-702-344-26	Sequence 26, Appl
C 36	37.2	10.4	240	1	US-08-628-417-6	Sequence 6, Appl
C 37	37.2	10.4	615	4	US-08-998-416-186	Sequence 186, App
C 38	37.2	10.4	837	4	US-08-998-416-1137	Sequence 1137, App
C 39	37.2	10.4	837	4	US-08-998-416-288	Sequence 288, App
C 40	37.2	10.4	860	1	US-07-847-010-18	Sequence 18, App
C 41	37.2	10.4	1615	1	US-08-413-118-108	Sequence 98, Appl
C 42	37.2	10.4	1615	2	US-08-413-118-108	Sequence 108, App
C 43	37.2	10.4	1615	2	US-08-473-446-108	Sequence 108, App
C 44	37.2	10.4	2058	2	US-08-749-391-1	Sequence 1, Appl
C 45	37.2	10.4	2058	3	US-09-390-200-1	Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
US-08-340-820-24/c  
Sequence 24, Application US/08340820  
Patent No. 5512460  
GENERAL INFORMATION:  
APPLICANT: NARUO, Ken-ichi  
APPLICANT: SEKO, Chisako  
APPLICANT: KUROKAWA, Tsutomu  
APPLICANT: KONDO, Tatsuya  
TITLE OF INVENTION: GLIA ACTIVATING FACTOR AND ITS  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS &  
STREET: 130 Water Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/340,820  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/835,713  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: CONLIN, David G.  
REGISTRATION NUMBER: 27026  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)523-3400  
TELEFAX: (617)523-6440  
TELEX: 200291 STRE UR  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1493 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens



Query Match	11.5%	Score 41	DB 4	Length 2230
Best Local Similarity	44.9%	Pred. No. 0.85		
Matches 155	Conservative 0	Mismatches 190	Indels 0	Gaps 0

  

OY	4	aaaaaagatbaaagatatttctatctctctcttttttttgataattttaa	catcatt	63
Db	2222	AACAAATATAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT		2163
OY	64	ttttggcccaatgatataataaatttggataataatataatttggatctt	ctttta	123
Db	2162	TTTTTTTTTAATGACGTAAAGACAAATTATGATGTCAAGTACATTT	TTTTTTTTTTTTTTTTTTTTTTTTTTTT	2103
OY	124	gttcgggtttgagaagaaggttctgcacttcgaaagtgcagatgatata	gatggag	183
Db	2102	TTTAACTTAACTTCATCTTTTGCAATTCATTAAATTAAGCTAAGACA	AAATTAATGATGTCAAGTA	2043
OY	184	ctaggcttgaagctcttggacattctgtatcgatgagctgttgatataa	gctgcac	243
Db	2042	CATTTTTTTTTTTTTTTTTTTTTTTTAACTTAACCTTCACCTTTGCA	TCATTAAATTAAGCTAAG	1983
OY	244	aaaccttaaatgggccttctataaaggcccaattatatacgatataa	caagaatgc	303
Db	1982	ACAATATATGATGTCTAAGTACATTTTTTTTTTTTTTTTTTTCATA	TAACTTCCTACT	1923
OY	304	tttactctgcttttggatccggaagcaataaacaattgtcaaatac		348

```

Db      1922 TTGGCATTCATTATAGCTAAGACAAATTATGATGTCTAAGTAC 1878
RESULT          5
US-09-056-075-1/c
; Sequence 1, Application US/09056075
; Patent No. 5953368
; GENERAL INFORMATION:
; APPLICANT: Johnson, Eric A.
; APPLICANT: Bradshaw, Marile
; APPLICANT: Rood, Julian
; TITLE OF INVENTION: Expression System for Clostridium
; TITLE OF INVENTION: Species
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Plinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056_075
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95238
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-9166
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6243 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 3770..4013
; OTHER INFORMATION: /note= "RP4 origin of DNA transfer (oriT) from
US-09-056-075-1 plasmid RP4"
```

Query Match	11.4%	Score 40.8;	DB 2;	Length 6243;
Best Local Similarity	47.3%;	Pred. No. 1;		
Matches 155;	Conservative 0;	Mismatches 172;	Indels 1;	Gaps 1;
QY 4	aaaaaagatgaagatattttatctctctcttttttggataattttaaactatt	63		
Db 1416	AAAAAATAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT	1357		
QY 64	tttttggccatgcatctaaaaatttggata-aaataattatggatctgtttt	122		
Db 1356	TTTTTATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT	1297		
QY 123	agtcggatttgaagaaagggttcgaacttcgaagtcgaagtcgatcatagatgsga	182		
Db 1296	TTTTTTTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT	1237		
QY 183	gcctagttgagctcttggacattgcatatggaagtgttgctatattgctgcacactat	242		
Db 1236	TTTTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT	1177		

Query Match	11.4%	Score 40.8;	DB 5;	Length 19557;
Best Local Similarity	52.3%	Pred. No. 1.1;		
Matches 90; Conservative	0;	Mismatches 82;	Indels 0;	Gaps 0;

[illegible]



```

; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-749-522-2

```

[illegible]

RESULT 11  
 US-08-341-568-3/c  
 Sequence 3, Application US/08341568  
 Patent No. 5661021  
 GENERAL INFORMATION:  
 APPLICANT: Bucherl, Johanna  
 APPLICANT: Silka-aho, Matti  
 APPLICANT: Viikari, Liisa  
 APPLICANT: Penttila, Merja  
 APPLICANT: Saloheimo, Anu  
 APPLICANT: Marjatta, Rannu  
 TITLE OF INVENTION: Mannanase enzymes, genes coding for them,  
 TITLE OF INVENTION: methods for isolating the genes, and methods for bleaching lignin  
 TITLE OF INVENTION: pulps  
 NUMBER OF SEQUENCES: 7  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Birch, Stewart, Kolasch and Birch  
 STREET: PO Box 747  
 CITY: Falls Church  
 STATE: VA  
 COUNTRY: US  
 ZIP: 22040-0747  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/341,568  
 FILING DATE: 22-NOV-1994  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Murphy Jr, Gerald M  
 REGISTRATION NUMBER: 28,977  
 REFERENCE/DOCKET NUMBER: 365-262P  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 205-8000  
 TELEFAX: (703) 205-8050  
 TELEX: 248345

```

: INFORMATION FOR SEQ ID NO: 3:
:
: SEQUENCE CHARACTERISTICS:
:
:   LENGTH: 289 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: single
:   TOPOLOGY: linear
:
: MOLECULE TYPE: cDNA
:
: ORIGINAL SOURCE:
:
:   ORGANISM: Trichoderma reesei
:   STRAIN: QM9414
:
:   US-08-341-568-3

```

	Query Match	11.0%;	Score 39.4;	DB 1;	Length 289;
	Best Local Similarity	67.9%;	Pred. No. 1.6;		
Matches	55; Conservative	0;	Mismatches	26;	Indels 0; Gaps 0.
QY	20 tatcttaacctcctcccttttcttgataaatltaaacatlttcctcgccaagtata	79			
Db	277 ttatgtgata	218			
QY	80 tataaaatlctggataaaataa	100			
Db	217 TATATATTATTCATCAAAA	197			

RESULT 12  
 US-08-911-020-3/c  
 : Sequence 3, Application US/08911020  
 : Patent No. 5854047  
 :  
 GENERAL INFORMATION:  
 APPLICANT: Buchert, Johanna  
 APPLICANT: Silka-aho, Matti  
 APPLICANT: Viikari, Liisa  
 APPLICANT: Penttila, Merja  
 APPLICANT: Saloheimo, Renu  
 APPLICANT: Marjatta, Renu  
 TITLE OF INVENTION: Mannanase enzymes, genes coding for them,  
 TITLE OF INVENTION: methods for isolating the genes, and methods for bleaching  
 TITLE OF INVENTION: lignocellulosic pulps  
 NUMBER OF SEQUENCES: 7  
 CORRESPONDENCE ADDRESSES:  
 ADDRESS: Birch, Stewart, Kolasch and Birch  
 STREET: PO Box 747  
 CITY: Falls Church  
 STATE: VA  
 COUNTRY: US  
 ZIP: 22040-0747  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/911.020  
 FILING DATE: 13-AUG-1997  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/341,568  
 FILING DATE: 22-NOV-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Murphy Jr, Gerald M  
 REGISTRATION NUMBER: 28,977  
 REFERENCE/DOCKET NUMBER: 365-262P  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 205-8000  
 TELEFAX: (703) 205-8050  
 TELEX: 248345  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 289 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single

```

65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
107
```









GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2001, 16:25:13 ; Search time 5433.22 Seconds  
(without alignments)  
622.857 Million cell updates/sec

Title: US-09-502-426-1\_COPY\_6111\_6468  
Perfect score: 358  
Sequence: 1 aaaaaaaaaagatgaagt.....tgtcaataccaacacag 358

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: gb\_est41:\*  
42: gb\_est42:\*  
43: gb\_est43:\*

44: em\_esthum10:\*  
45: em\_esthum11:\*  
46: em\_esthum12:\*  
47: em\_esthum13:\*  
48: em\_esthum14:\*  
49: em\_esthum15:\*  
50: em\_esthum16:\*  
51: em\_esthum17:\*  
52: em\_esthum18:\*  
53: em\_esthum19:\*  
54: em\_esthum20:\*  
55: em\_esthum21:\*  
56: em\_esthum22:\*  
57: em\_esthum23:\*  
58: em\_esthum24:\*  
59: em\_esthum25:\*  
60: em\_esthum26:\*  
61: em\_esthum27:\*  
62: em\_esthum28:\*  
63: em\_esthum29:\*  
64: em\_esthum30:\*  
65: em\_esthum31:\*  
66: em\_esthum32:\*  
67: em\_esthum33:\*  
68: em\_esthum34:\*  
69: em\_esthum35:\*  
70: em\_esthum36:\*  
71: em\_esthum37:\*  
72: em\_esthum38:\*  
73: em\_esthum39:\*  
74: em\_esthum40:\*  
75: em\_esthum41:\*  
76: em\_esthum42:\*  
77: em\_esthum43:\*  
78: em\_esthum44:\*  
79: em\_esthum45:\*  
80: em\_esthum46:\*  
81: em\_esthum47:\*  
82: em\_esthum48:\*  
83: em\_esthum49:\*  
84: em\_esthum50:\*  
85: em\_esthum51:\*  
86: em\_esthum52:\*  
87: em\_esthum53:\*  
88: em\_esthum54:\*  
89: em\_esthum55:\*  
90: em\_esthum56:\*  
91: em\_esthum57:\*  
92: em\_esthum58:\*  
93: em\_esthum59:\*  
94: em\_esthum60:\*  
95: em\_esthum61:\*  
96: em\_esthum62:\*  
97: em\_esthum63:\*  
98: em\_esthum64:\*  
99: em\_esthum65:\*  
100: em\_esthum66:\*  
101: em\_esthum67:\*  
102: em\_esthum68:\*  
103: em\_esthum69:\*  
104: em\_esthum70:\*  
105: em\_esthum71:\*  
106: em\_esthum72:\*  
107: em\_esthum73:\*  
108: em\_esthum74:\*  
109: em\_esthum75:\*  
110: em\_esthum76:\*  
111: em\_esthum77:\*  
112: em\_esthum78:\*  
113: em\_esthum79:\*  
114: em\_esthum80:\*  
115: em\_esthum81:\*  
116: em\_esthum82:\*

```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

```

```

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.





<b>FEATURES</b>					
Source	Location/Qualifiers				
	1. .1101				
	/organism="Drosophila melanogaster"				
	/plasmid="pbeloBAC11"				
	/db_xref="taxon:7227"				
	/clone_lib="DrosBAC"				
	/clone="BACN16D22"				
	/note="End : 177"				
BASE COUNT	203 a	220 c	84 g	158 t	436 others
ORIGIN					
Query Match	16.2%	Score 58:	DB 219:	Length 1101:	
Best Local Similarity	34.3%	Pred. No.	0.097:		
Matches 122:	Conservative	76:	Mismatches 156:	Indels 2:	Gaps 1:
OY	1 aaaaaaaaaaaagataaagatattcttccttttcttttttttgataatttaacaa 60				
Db	737 WMAAAAAATTTATDTTTTTTTTTTWTWTTTBTTTAATMTTAKDDTTTTTTTW 796				
OY	61 ttctttgcccaatgatataaaaaatttgtataaatattatgtatatcgtttt 120				
Db	797 AATATTWTWTWDATWTTTTTTTTTTTWMATATATTTTMMTAAAT--AMATWMTATT 854				
OY	121 ttagtcggggtttcgagaagggttcgcacttcgaagtcgaagtcgacgatagatctgg 180				
Db	855 AMTATAMWATATMTWWMMWRRTAMWTDWTMMWMTAKTKTMTATWATATWAMWMMWA 914				
OY	181 gsgctaggttgagcttcgttgacaattgttatgtgatgttgatattagtgcgaact 240				
Db	915 MAMWAMAATTATWMTWMTATTADMTPTTASTATWTTTDPWTATWARAKARWATWDMAN 974				
OY	241 attaacctctaattgagcttcctatataagcccaattatataagattacgataacaaagtcac 300				
Db	975 AMRATMDTATWATAADATWATATAAAMWMTWDTATTGCRMAMTADATTAMDAAATTRA 1034				
OY	301 aaecttacttcglttttgatccgaagcaataaacaattgtccaataaccacaaca 356				
Db	1035 AATWTTTWTMTWKMTTKWTTTKTKAKTKAGRAMWWWRTKAARRWMTTRADA 1090				
RESULT 6	AZ691838	961 bp	DNA	GSS	14-DEC-2000
AZ691838/c					
LOCUS	ENTM156F Entamoeba histolytica Sheared DNA Entamoeba histolytica				
DEFINITION	genomic, DNA sequence.				
ACCESSION	AZ691838				
VERSION	AZ691838.1	GI:11828984			
KEYWORDS	GSS.				
SOURCE	Entamoeba histolytica.				
ORGANISM	Entamoeba histolytica				
REFERENCE	Eukaryota; Eumetazoa; Excavates; Amoebozoa; Entamoebidae; Entamoeba.				
AUTHORS	1 (bases 1 to 961)				
TITLE	Loftus,B., Van Aken,S. and Fraser,C. Determination of clone end sequences from Entamoeba histolytica HM1:IMSS sheared DNA library				
JOURNAL COMMENT	Unpublished (2000) Contact: Brendan J Loftus Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: bjo@fsl.isr.org Clones are derived from the Entamoeba histolytica HM1:IMSS sheared DNA library Seq primer: MJ3-Forward Class: Shotgun				





[illegible]

ORIGIN	15.8%; Score 56.4; DB 219; Length 1101;
Query Match	Best Local Similarity 47.4%; Pred. No. 0.19;
Matches 83; Conservative 24; Mismatches 68; Indels 0; Gaps 0;	
OY 2	2 aaaaaaaagatgaagatatttattctctctcttttttttgataatttaacat 61
Db 892	WAAAAAAMAAAAAAMWTTTTTTTTTAAATTTTMTTAAATTTTAAATAAAAA 951
OY 62	ttttttgccatgatataataatttgataataatattatgtatattcgttttt 121
Db 932	TTTTTGMATTTWTTTNTNDAGTAAMAAATTTTMTTAAATTTTTRTTTTTTT 1011
OY 122	taagtgcggattgagaaaagtttcagacttcgaaattgagcagatagata 176
Db 1012	TTTTTTTTTATATGACATRTTTTTTTTTTTTAAAGAGATTADAAATTTAA 1066
RESULT 10	
CNS0021J/c	1101 bp DNA GSS 03-JUN-1999
LOCUS	
DEFINITION	CNS0021J melanogaster genome survey sequence TET3 end of BAC #
ACCESSION	BACR05N11 of RPCI-98 library from Drosophila melanogaster (fruit
VERSION	fly), genomic survey sequence.
KEYWORDS	AL061936
SOURCE	AL061936.1 GI:4940214
ORGANISM	GSS.
REFERENCE	fruit fly.
AUTHORS	Drosophila melanogaster
JOURNAL	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
	1 (bases 1 to 1101)
COMMENT	Genoscope.
	Direct Submission
	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
	BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
	- Web : www.genoscope.cns.fr)
	Determination of this BAC-end sequence was carried out as part of a
	collaboration with the Berkeley Drosophila Genome Project (BDGP).
	The BDGP is constructing a physical map of the Drosophila
	melanogaster genome using these BACs. For further information
	please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila
	melanogaster BAC library was prepared by Kazuhiro Osoegawa and
	Aaron Mammosser in Pieter de Jong's laboratory in the Department of
	Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
	NY. The library is named RPCI-98 and was constructed by partial
	ECORI digestion of Drosophila DNA provided by the BDGP from the
	isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
	PL and EST libraries. A more detailed description of the library
	and how to order individual BAC clones, the entire library, or
	filters for hybridization from the BACPAC Resource Center can be
	found at <a href="http://bacpac.med.buffalo.edu/drosophila_bac.htm">http://bacpac.med.buffalo.edu/drosophila_bac.htm</a> .
FEATURES	location/Qualifiers
source	1..1101
	/organism="Drosophila melanogaster"
	/db_xref="taxon:7227"
	/clone_lib="RPCI-98"
	/clone="BACR05N11"
	/note="end : TET3"
BASE COUNT	631 a 7 c 28 g 289 t 146 others
ORIGIN	
Query Match	15.7%; Score 56.2; DB 219; Length 1101;
Best Local Similarity 43.3%; Pred. No. 0.21;	
Matches 153; Conservative 28; Mismatches 171; Indels 1; Gaps 1;	
OY 1	aaaaaaaaaagatgaagatatttattctctctcttttttttgataatttaac 60
Db 873	ATAAATAATATTTTAAAAAAMWTTTTTTTTTAAATTTTAAATTTTATTTT 814

[illegible]

RESULT	11
CNSO4AIH/C	Locus
CNSO4AIH	949 bp DNA GSS
Tetradon nigroviridis genome survey sequence T7 end of clone 095F19 of library G from Tetradon nigroviridis, genomic survey sequence.	21-MAY-2000
AL281906	
ACCESSION	AT281906.1 GI:8020236
VERSION	GSS; genome survey sequence.
KEYWORDS	Tetradon nigroviridis..
SOURCE	Tetradon nigroviridis
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorphi; Tetraodontiformes; Tetraodonidae; Tetraodon. 1 (bases 1 to 949)
REFERENCE	Roeest-Crolius,H., Jalllon,O., Dasliya,C., Fizames,C., Fisher,C., Bonneau,L., Billault,A., Quetler,F., Saurin,W., Bernot,A. and Weissenbach,J.
AUTHORS	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradon nigroviridis
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 949)
AUTHORS	Roeest-Crolius,H., Jalllon,O., Dasliya,C., Bonneau,L., Fisher,C., Bernot,A., Fizames,C., Winkler,P., Brottier,P., Quetler,F., Saurin,W. and Weissenbach,J.
TITLE	Human gene number estimate provided by genome wide analysis using Tetradon nigroviridis DNA sequence
JOURNAL	Unpublished
REFERENCE	3 (bases 1 to 949)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL COMMENT	Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases This sequence is a single read and was generated as part of a large scale clone-and-sequencing project of the Tetradon nigroviridis genome. For more information, please take a look at <a href="http://www.genoscope.cns.fr/Tetraodon">http://www.genoscope.cns.fr/Tetraodon</a> .

```

FEATURES
source
    location/Qualifiers
    1..949
    /organism="Tetrarodon nigroviridis"
    /db_xref="taxon:99883"
    /clone="095F19"
    /clone_11b="G"
    /note="Genoscope sequence ID : C0BG09SCC10LPI-end : T7"
BASE COUNT
    355 a 144 c 185 g 223 t 42 others
ORIGIN

```

Query Match	15.6%;	Score 56;	DB 221;	Length 949;
Best Local Similarity	49.3%;	Pred. No. 0.23;		
Matches 74;	Conservative 21;	Mismatches 55;	Indels 0;	Gaps 0;

```

QY      4  aaaaaagaagaaagatcttttaccctctcttttttctgtaattcttaacatct 63
      820  :::::||||| : ::::|:||||| : : |||
      820  WMAAAAAAAAAAAMWTTTTTTTTTTTTTTTTTTTTTTTTTTTTNNNNNAATTTTTTT
Db
QY      64  ttctgcaccaatgataataaaattcgagaaataatattggaatattcgtcttca 122
      ||||| :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Dd      760  TTTTTTTTTTMMWTTTTTMMMAAMTTTTTTTTTTTTTTTTTTTTMMNNMMWTTTTAAAMTTTAAATTTTTT 701
QY      124  gtctcggtttgagaaaagggttcgcactt 153
      || ||| ||| : |||
Dd      700  TTTTTTTTTTTTAAAMTTTTTTTTMMWTTT 671

```

RESULT	12
CNS000AQX/c	
LOCUS	CNS000AQX 767 bp DNA GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR21A20 of Rpci-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL055924
VERSION	AL055924.1 GI:4936693
KEYWORDS	GSS.
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 767)

REFERENCE AUTHORS TITLE JOURNAL	COMMENT
1 (bases 1 to 767) Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr	Web : <a href="http://www.genoscope.cns.fr">www.genoscope.cns.fr</a> Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <a href="http://www.fruitfly.org/TheBDGP/Drosophila">http://www.fruitfly.org/TheBDGP/Drosophila</a>
	Melanogaster BAC library was prepared by Kazutoyo Oseegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2: cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource center can be found at <a href="http://bacpac.med.buffalo.edu/drosophila_bac.htm">http://bacpac.med.buffalo.edu/drosophila_bac.htm</a> .

FEATURES	source	location/Qualifiers
1.	.767	/organism="Drosophila melanogaster"
		/db_xref="taxon:7227"
		/clone_lib="RPCI-98"
		/clone="BACR21A20"
		/note="end : TER3"
BASE COUNT	312 a	105 c 65 g 164 t 121 others
ORIGIN		

Query Match	15.6%;	Score 55.8;	DB 219;	Length 767;
Best Local Similarity	33.1%;	Pred. NO. 0.25;		
Matches	99;	Conservative	64;	Mismatches 136;
			Indels	0;
			Gaps	0;

Oy	23	tttatctctctctcttcttcttgataaatcttaacatcttcttcccaapabat	82
	: :         : :     : :     : :     :		
Dd	389	TATTTMMTTTTTTTTTGGTTTTMMWMTOKMWTMTMTTKTKMKAKKTTTTTTT	330
Oy	83	aaaaatttgtataatatatattgtacatctgctttcagtcgggtcttggaagaag	142
	: :     : :     : :     : :     : :     :		
Dd	329	TTTGWGAATAKKMMKTAAKKMKTTMTTTTTTTTTTTTTTTTATMAAAGAMMAAK	270
Oy	143	gttcgcacttcgaagaagtgcagatgatatabagattvggagctaggcttgacctggac	202

Qy	Db	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572
----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

[illegible]





PS Claim 50; Fig 11; 113pp; English.

XX The present sequence represents a DMF4 polypeptide. The polypeptide is a

CC cytochrome P450 enzyme that mediates multiple steps in synthesis of

CC brassinosteroids. Specifically, it mediates multiple

CC 22alpha-hydroxylation steps in brassinosteroid biosynthesis. The DMF4

CC polynucleotide is used for altering the phenotype of a plant. DMF4

CC plants display a dramatic reduction in the length of different organs,

CC and this size reduction is attributable to a defect in cell elongation.

CC The DMF4 polynucleotides and polypeptides can be used in diagnostic

CC assays and to generate antibodies, which can be used to produce

CC immunogenic compositions.

XX

XX Sequence 513 AA:

SO

Query Match 100.0%; Score 2681; DB 21; Length 513;

Best Local Similarity 100.0%; Pred. No. 6,6e-229;

Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFEHTHTLLPILLPSLLLEFILLKRRNRKTRFNLPGKSGMPFGTIGYLPY 60

DB 1 mfehthtllp illpslllefillkrrnrktrfnlpgksgmpf g t i g y l p y t 60

QY 61 ATTGDEMOQHVSKYKTYRSNLFGEPTIVSADAGLNRFILQNEGRLEFECYSPTGIL 120

DB 61 attgdfmqhvkyskyktyrsnlfg eptivs adaglnrfilqnegrlfecysptg il 120

QY 121 GKWSMLYVGMHDMRSISNPLSHARLRTILKDVGRHRLFLVLDSDMQNSISADQEA 180

DB 121 gkwsml yv g m h d m r s i s n p l s h a r l r t i l k d v g r h r l f l v l d s d m q n s i s a d q e a 180

QY 181 KKFENLAKIMSMDSPEEETQLEKKEYVTFMKGVSAAPNLPGTAVHKALOSRAITLK 240

DB 181 k k f e n l a k i m s m d s p e e e t q l e k k e y v t f m k g v s a a p n l p g t a v h k a l o s r a i t l k 240

QY 241 FIERKMERKLDKEDEEVEKTEDEAKMSKSDHVRKQRTDDLLGVLKHSNLSTEQ 300

DB 241 f i e r k m e r k l d k e d e e v e k t e d e a k m s k s d h v r k q r t d d l l g v l k h s n l s t e q 300

QY 301 ILDLILSLFSGHETSSVAIALAFPLQACPKAYELREHLELAKAKKEGSESLMDD 360

DB 301 i l d l i l s l f s g h e t s s v a i a l a f p l q a c p k a y e l r e h l e l a k a k k e g s e s l m d d 360

QY 361 YKKMDFQCVINETLRIGNVVRFRLRKALKDVRKGYDIPSGMKVLPVISAHLNDSRYD 420

DB 361 y k k m d f q c v i n e t l r i g n v v r f l r k a l k d v r k g y d i p s g m k v l p v i s a h l n d s r y d 420

QY 421 QPULFNPWRMOQONNGASSSGSGSFTWGNMYMPFGGPRLCAGSELAKLEMAVFIHHLV 480

DB 421 q p u l f n p w r m o q o n n g a s s s g s f t w g n m y m p f g g p r l c a g s e l a k l e m a v f i h h l v 480

QY 481 LKFMWELAEEDDPRAPFVDPNGLPIRVSRIL 513

DB 481 l k f m w e l a e e d d p r a p f v d p n g l p i r v s r i l 513

RESULT 2

AAW27153

ID AAW27153 standard; Protein: 472 AA.

XX

AC AAW27153;

XX

DT 14-Apr-1998 (first entry)

XX

DE Arabidopsis thaliana cytochrome P450-type hydroxylase.

XX

KM Cytochrome P450-type hydroxylase; identification; brassinosteroid;

KM brassinosteroid inhibitor; modified plant; recombinant production;

KM teasterone.

XX

OS Arabidopsis thaliana.

XX

PN W09735986-A1.

XX

PD 02-Oct-1997.

XX

PF 27-MAR-1997; 97WO-EP01586.

XX

PR 27-MAR-1996; 96US-0622166.

XX

PA (PLAC) MAX PLANCK GES FORDERUNG WISSENSCHAFTEN.

XX

PI Altmann T, Koncz C, Mathur J, Szekeres MA;

XX

XX WPI: 1997-489649/45.

DR

DR N-PSDB: AAT85306, AAT85307.

XX

PT New isolated plant cytochrome P450-type hydroxylase gene - used to

PT identify substances acting as brassino-steroid(s) or brassinosteroid

PT inhibitors for the production of modified plants

XX

PS Claim 1; Pages 44-46; 77pp; English.

XX

CC The present sequence is Arabidopsis thaliana cytochrome

CC P450-type hydroxylase. The hydroxylase can be used to identify

CC brassinosteroids or brassinosteroid inhibitors, useful to produce

CC plants with modified physiological and/or phenotypic

CC characteristics. The modified plants may show, e.g. stimulation of

CC growth, increased cell elongation, increased wood production,

CC accelerated seed germination at low temperatures, an increase in

CC dry weight, repressed anthocyanin production during growth in light

CC and/or inhibited de-etiolation which is induced, e.g. by cytokinin,

CC in the dark or an increase in stress tolerance. The hydroxylase or

CC its coding sequence can also be used for the recombinant production

CC of compounds, e.g. teasterone.

XX

SO Sequence 472 AA;

Query Match 36.5%; Score 978.5; DB 18; Length 472;

Best Local Similarity 41.9%; Pred. No. 3e-78;

Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

QY 12 LLLPSSLLEFILLKRRNRKTRFNLPGKSGMPFGTIGYLPYATVATLQDMQOH 71

DB 12 l l l p s l l e f i l l k r r n r k t r f n l p g k s g m p f g t i g y l p y a t v a t l q d m q o h 71

QY 72 VSKYKTYRSNLFGEPTIVSADAGLNRFILQNEGRLEFECYSPTGILCKWSMLVGYD 131

DB 72 v s k y k t y r s n l f g e p t i v s a d a g l n r f i l q n e g r l e f e c y s p t g i l c k w s m l v g y d 131

QY 132 MHRDMRSISNPLSHARLRTILKDVGRHRLFLVLDSDMQNSIFSAODEAKKFTFNMAKH 191

DB 132 m h r d m r s i s n p l s h a r l r t i l k d v g r h r l f l v l d s d m q n s i f s a o d e a k k f t f n m a k h 191

QY 192 IMSDPGEETQLEKKEYVTFMKGVSAAPNLPGTAVHKALOSRAITLKETERMEERKL 251

DB 192 i m s d p g e e t q l e k k e y v t f m k g v s a a p n l p g t a v h k a l o s r a i t l k e t e r m e e r k l 251

QY 252 DIKEDDEEVEKTEDEAKMSKSDHVRKQRTDDLLGVLKHSNLSTEQILLLSLLA 311

DB 252 d i k e d d e e v e k t e d e a k m s k s d h v r k q r t d d l l g v l k h s n l s t e q i l l l s l l a 311

QY 312 GHETSSVAIALAFPLQACPKAYELREHLELAKAKKEGSESLMDDYKKMDFQCVI 371

DB 312 g h e t s s v a i a l a f p l q a c p k a y e l r e h l e l a k a k k e g s e s l m d d y k k m d f q c v i 371

QY 372 NETLRIGNVVRFRLRKALKDVRKGYDIPSGMKVLPVISAHLNDSRYDQPNLFNPRMQ 431

DB 372 n e t l r i g n v v r f l r k a l k d v r k g y d i p s g m k v l p v i s a h l n d s r y d q p n l f n p r m q 431

QY 432 QONNGASSSGSFTWGNMYMPFGGPRLCAGSELAKLEMAVFIHHLVLFMWELEAD 490

DB 432 q o n n g a s s s g s f t w g n m y m p f g g p r l c a g s e l a k l e m a v f i h h l v l f m w e l e a d 490

QY 490

DB 490

OY 491 DQPFAPFVDFPNCGLPIRVSR 511  
|: || || |  
Db 447 dklvlfpttrtqkryplfvkr 467

## RESULT 3

AAG44571

ID AAG44571 standard; Protein: 472 AA.

XX AAG44571;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 55847.

XX Protein identification: signal transduction pathway; metabolic pathway;  
XX hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.

XX Arabidopsis thaliana.

XX EPI033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127452.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

XX 19-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130449.

XX 23-APR-1999; 99US-0130510.

XX 23-APR-1999; 99US-0130891.

XX 28-APR-1999; 99US-0131449.

XX 30-APR-1999; 99US-0132048.

XX 04-MAY-1999; 99US-0132407.

XX 05-MAY-1999; 99US-0132484.

XX 06-MAY-1999; 99US-0132485.

XX 06-MAY-1999; 99US-0132486.

XX 07-MAY-1999; 99US-0132487.

XX 11-MAY-1999; 99US-0132863.

XX 14-MAY-1999; 99US-0134256.

XX 14-MAY-1999; 99US-0134218.

XX 14-MAY-1999; 99US-0134219.

XX 14-MAY-1999; 99US-0134221.

XX 14-MAY-1999; 99US-0134370.

XX 18-MAY-1999; 99US-0134768.

XX 19-MAY-1999; 99US-0134941.

XX 20-MAY-1999; 99US-0135124.

XX 21-MAY-1999; 99US-0135353.

XX 24-MAY-1999; 99US-0135629.

XX 25-MAY-1999; 99US-0136021.

XX 27-MAY-1999; 99US-0136392.

XX 28-MAY-1999; 99US-0136782.

XX 01-JUN-1999; 99US-0137222.

XX 03-JUN-1999; 99US-0137528.

XX 04-JUN-1999; 99US-0137502.

XX 07-JUN-1999; 99US-0137724.

XX 08-JUN-1999; 99US-0138094.

XX 10-JUN-1999; 99US-0138540.

XX 10-JUN-1999; 99US-0138847.

XX 14-JUN-1999; 99US-0139119.

XX 16-JUN-1999; 99US-0139452.

XX 16-JUN-1999; 99US-0139453.

XX 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139889.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147933.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.

PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151055.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152353.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 26-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 36.5% Score 978.5; DB 21; Length 472;  
Best Local Similarity 41.9%; Pred. No. 3e-78;  
Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

QY 12 LLLPSTLLFLILKRRNRKTRFNLPGKSGMPFLGFTIGYLVKPYTATTLDGFMQOH 71  
DB 7 LLLLSlaagfl--lllrtrrmgippslpllgelcfqlgykltenepepflder 63  
QY 72 VSKTGKIRSLGEGPTIVSADAGLNRFILONEGRLEFCSIPRSIGILGKWSMLVLVGD 131  
DB 64 varygsvfmlthfgeptlfsadepctnrflvqneqklfecsyasncilghsillmkgs 123

QY 132 MHRDRSISLNFSLHARLRTILKDKVERHTLEVLDSMOONSIFSADBEAKKFTFNLMAKH 191  
DB 124 lhrkthslmsfiansslikdhlmdidtlvfnldswsryll--meaakkltelcvkq 181  
QY 192 IMSMDPGEEETEOLEKKEVYTFMKGVSAPLNLPSTAYHKALOSATTLKFTERRMEERKL 251  
DB 182 lmsfdpg-ewseslrkeylllviegffsirlplfsttykqldar-----rkvaalt 232  
QY 252 DIKEEDDEEEVKTEDEKEMSKSDHVRKQRTDDLLGKSNVLSHNSLSTFOILLLSLFLA 311  
DB 233 vvmkrrreeeeegae-----rkkdmllaallaaaddgfsdeevdfivaliva 278  
QY 312 GHETSSVALAIAIFLQACPKAVELEREHLEIRAKKELESESLNMDYKKMPTOCVI 371  
DB 279 gyetstlmtlavflicetplalaglkhehekiramksd--syalewsdyksmpftcgv 336  
QY 372 NETLRIGNVRFELHRAKALKDVRKGYDIPSGKVLVYISAVHLDNSRYDQNLFPWRMQ 431  
DB 337 netlrvaalligvframtdveikgkirkqwkvfssfravhldpnhkdarftnprtwg 396  
QY 432 QONNGASSSGSFSSTWGN--YMPFGGGRCLACAGSELAKLEMAVFIIHLVLFKFWELAED 490  
DB 397 -----snvltgpsnvftfpfg99prlcpylatavalavflnrlvgtfswpaeg 446  
QY 491 DQPAFPFVDFPNGLPIRVS 511  
DB 447 dkivfptrtqrkryplfvkr 467  
  
RESULT 4  
AAG45022  
ID AAG45022 standard; Protein: 472 AA.  
AC AAG45022;  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 56469.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
OS Arabidopsis thaliana.  
PN EP1033405-A2.  
XX  
XX 06-SEP-2000.  
PD  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.



PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139817.  
PR 23-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 15-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147935.  
PR 09-AUG-1999; 99US-0148177.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151308.  
PR 01-SEP-1999; 99US-0151330.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157113.  
PR 05-OCT-1999; 99US-0157177.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.

PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 36.5% Score 978.5; DB 21; Length 472;  
Best Local Similarity 41.9%; Pred. No. 3e-78;  
Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

QY 12 LLLPSLTLFLILKRRNRKTRFNLPGKSGWPLGFTTGYLKPYPATTLGDFMOOH 71  
Db 7 LLLSLAAGFL---LIRIRYRMGLPPGSLGPIPLIGETFGIAGYKktenpepfider 63  
QY 72 VSKGKIRKSNLFEPTIVSADAGINRFLQNEGRLEPCSTPRSTIGILGKWSMLVLVD 131  
Db 64 VARYSVMTLHfgepltfdsadpelnrfvlqneqklfecsyasacnllgkhslllmkgs 123  
QY 132 MHRDRSISLNLFSHARLRTILKOVERHTLFLVDSMQONSIFSAODRAKKETFLMAKH 191  
Db 124 Lhkrmhsilmsfmsalikhldiditlvrfnldswsrrvll--meaakkltfelvkvq 181  
QY 192 IMSMDPGEETEOLEKREYVTFMGVSNAPNLPGTAHKALOSRAITILKFERKMEERKL 251  
Db 182 lmsfdpg-eweselsrkeyllvlegffslpllfstlykaigar-----rkvaalt 232  
QY 252 DIKEEDDEEEVKEDEKEMASKSDHVRKQRTDILLGVNLKNSLSTQIIDLILSLFA 311  
Db 233 vvmkrrteeegae-----rkkdmllaadgfsdeelvdlvalla 278  
QY 312 GHETSSVAIALAIFELQCPKAVEELREHLEIARAKKEGSESLNMDYKKMPTOCVI 371  
Db 279 gyetstlmtlavrlfctplalaglkcheekiramsd--syslewsdyssmptlqcv 336  
QY 372 NETRLGNVRFELRRKALKDVRKGYDIPSGKYLPLVISAHLNDSRYDQNLFPNWMQ 431  
Db 337 netlrvalilgvrfrantdvleikykikpkykwfssfravhldpnhkdaartlfpwrq 396  
QY 432 QONNCASSSGSFTWCNN--YMPGCGPRLCAGSESLAKLEMAVFIHHLVLFKFMELAE 490  
Db 397 -----snavltgpsnvfcpfpggprlcpgyelarvalsvflhrlvlgfswpaeg 446  
QY 491 DQPAFPVDFPNCPLIRVSR 511  
Db 447 dklvffpttrqtkypflvkr 467

RESULT 5  
AAG45021  
ID AAG45021 standard; Protein: 491 AA.

XX AAG45021;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 56468.

KW Protein identification: signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX Arabidopsis thaliana.

XX EPI033405-A2.

XX

PD 06-SEP-2000.  
XX  
PE 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126284.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.

```

PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.

```

```

PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156506.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158368.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 36.5%; Score 978.5; DB 21; Length 491;  
Best Local Similarity 41.9%; Pred. No. 3.2e-78;

Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

```

QY 12 LLLPSSLLFLFLRRNRKTRFNLPGCKSGMPFGETIGYLKPYATTLGSPMOOH 71
DB 26 LLLSSlaagl--llrrtyrmglppsglglplgetlqlgayktenpeptlver 82
QY 72 VSKYKIVRSNLFCEPTIVSADAGLNRFILONEGRLFECSSYPSISGILGKSMVLVGD 131
DB 83 varygsvfmtthlfgpeptlfsadpctnrfvlgnegrlfecsyfasicnllgkshllmlygs 142
QY 132 MHRDKRSISLNFSLHARLRTILKDKVERHTLFLVDSMOONSIFSAODEAKKFTPLMAKH 191
DB 143 lkhmrhslmsfmsfssllkdhlmldidrtvfnldwsarvll--meeakktlftlvkq 200
QY 192 IMSNDPGEETEQLKKEYVTFMKGVSAPLNPGAYHKALOSRAITILKTRKKMEERL 251
DB 201 lmsfdpg-ewseslkeyllvleglfsplrlfstlyraiaqar-----rkvaeatl 251
QY 252 DIKEDEEEEVKTEDEEMSKSDHVVRKQRTDDDLGWLKHSNLTSEQILDLLSLFLA 311
DB 253 vvmvkmrreeeegeae-----rkkmldlaallaaddgsdeevlfiavallva 297
QY 312 GHETSSVAIALAIFFLQACPVAEELREHEIARAKKELGSELSNMDYKKMDFTQCVI 371
DB 298 gyetstlmtlavkflteqplaqkleehekltamksd--syslswsykxmpftgcgv 355
QY 372 NETLRIGNVVRFLHRRKALKDVKYKGYDTPSGKMYLPLVSAVILDSRVDQPLFLPWRWQ 431
DB 356 netlvvanlvgvtramdveikykrlpkvkvfssfravlnlphnhdartttrpwq 415
QY 432 QONNGASSSGSFSFTWGN--YMPFGGPRLCAGSELAKLEMAVFIHHLVLKFMNELAED 490

```

Db 416 -----snsvltspnvtfpfg99prrlcpylarvalsvflhrltvgfswpaed 465  
Oy 491 DOPFAFPVDFPNGPIRVSR 511  
|:| | | |  
Db 466 dklvfpctrtqkryplvkr 486

RESULT 6  
AAG44570  
ID AAG44570 standard; Protein: 492 AA.

XX AAG44570;  
AC  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 55846.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PE 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139859.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147403.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565;  
PR 13-AUG-1999; 99US-0148684;  
PR 16-AUG-1999; 99US-0149368;  
PR 17-AUG-1999; 99US-0149175;  
PR 18-AUG-1999; 99US-0149426;  
PR 20-AUG-1999; 99US-0149722;  
PR 20-AUG-1999; 99US-0149723;  
PR 20-AUG-1999; 99US-0149929;  
PR 23-AUG-1999; 99US-0149920;  
PR 23-AUG-1999; 99US-0149930;  
PR 25-AUG-1999; 99US-0150566;  
PR 26-AUG-1999; 99US-0150884;  
PR 27-AUG-1999; 99US-0151065;  
PR 27-AUG-1999; 99US-0151066;  
PR 27-AUG-1999; 99US-0151080;  
PR 30-AUG-1999; 99US-0151303;  
PR 31-AUG-1999; 99US-0151438;  
PR 01-SEP-1999; 99US-0151930;  
PR 07-SEP-1999; 99US-0152363;  
PR 10-SEP-1999; 99US-0153070;  
PR 13-SEP-1999; 99US-0153758;  
PR 15-SEP-1999; 99US-0154018;  
PR 16-SEP-1999; 99US-0154039;  
PR 20-SEP-1999; 99US-0154779;  
PR 22-SEP-1999; 99US-0155139;  
PR 23-SEP-1999; 99US-0155486;  
PR 24-SEP-1999; 99US-0155659;  
PR 28-SEP-1999; 99US-0156458;  
PR 29-SEP-1999; 99US-0156596;  
PR 04-OCT-1999; 99US-0157117;  
PR 05-OCT-1999; 99US-0157753;  
PR 06-OCT-1999; 99US-0157865;  
PR 07-OCT-1999; 99US-0158029;  
PR 08-OCT-1999; 99US-0158232;  
PR 12-OCT-1999; 99US-0158369;  
PR 13-OCT-1999; 99US-0159293;  
PR 13-OCT-1999; 99US-0159294;  
PR 13-OCT-1999; 99US-0159295;  
PR 14-OCT-1999; 99US-0159329;  
PR 14-OCT-1999; 99US-0159330;  
PR 14-OCT-1999; 99US-0159331;  
PR 14-OCT-1999; 99US-0159637;  
PR 14-OCT-1999; 99US-0159638;  
PR 14-OCT-1999; 99US-0159584;  
PR 16-OCT-1999; 99US-0160770;  
PR 21-OCT-1999; 99US-0160814;  
PR 21-OCT-1999; 99US-0160815;  
PR 21-OCT-1999; 99US-0160980;  
PR 22-OCT-1999; 99US-0160981;  
PR 22-OCT-1999; 99US-0160989;  
PR 22-OCT-1999; 99US-0161404;  
PR 25-OCT-1999; 99US-0161405;  
PR 25-OCT-1999; 99US-0161406;  
PR 25-OCT-1999; 99US-0161359;  
PR 26-OCT-1999; 99US-0161360;  
PR 26-OCT-1999; 99US-0161361;  
PR 28-OCT-1999; 99US-0161920;  
PR 28-OCT-1999; 99US-0161992;  
PR 28-OCT-1999; 99US-0161993;  
PR 29-OCT-1999; 99US-0162142;  
Query Match 36.5%; Score 978.5; DB 21; Length 492;  
Best Local Similarity 41.9%; Pred. No. 3.2e-76;  
Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

Db 84 vaigsvfemthlfgpeptlfsadpeneftvlqnegklfscspaaicnllghkslllmgs 143  
Qy 132 MHRMRSISLNFSLHARLRTLLKDVREHRTLFDVDSWOONSIFSADAKKFTENLMAKH 191  
Db 144 lhrkhsilmsfamsilkhlmldidrlvrfnldswesrvll--meaakkltfelvkq 201  
Qy 192 IMSMDGEEETOLKKEVVTMKGVSAPLNPGVAHKAQSRATILKFERKKEEKL 251  
Db 202 lmsitpg-eweserkeyllvlegflslpflstlytkaiqar-----rkvaalt 252  
Qy 252 DIKEEDQEEFEVKTEDEAFEMSKSDHVRKQRTDDDLGWLKHSNSTCOIDLJLSLFA 311  
Db 253 vvmkrrreeeegea-----rkmdaalaaaddgsdeideivdvaliva 298  
Qy 312 GHETSSVALALAIFFLQACPRAVEELREHLEIARAKKELGSELNWDYKKMDFTCVI 371  
Db 299 gyetstlmtlavkfltetplalaqlkeehetiramed--sylewedyksmptqcvv 356  
Qy 372 NETLRIGNVVRFILHRKALKDVRKGYDIPSGMKVLPVISAHLNDSRYDQPLFPMRMQ 431  
Db 357 netlrvanllgvtfrantdveikgkypkwkvfssfravhlphnkdartlptwq 416  
Qy 432 QONNGASSSGSFSFTWGN--YMPGCGPRLCAGSELAKEVAFTIHLVLKFNELAE 490  
Db 417 -----snvltgspnvtftpggprlcpgeylarvalsvllhvlvgfswpaeg 466  
Qy 491 DQFAFPFVDEPNGLPIRVSR 511  
Db 467 dklvffptctckryplfvkr 487

RESULT 7  
AAG4572  
ID AAG4572 standard; Protein: 444 AA.  
XX  
AC AAG4572;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 55848.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX  
OS Arabidopsis thaliana.  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PE 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825;  
PR 05-MAR-1999; 99US-0123180;  
PR 09-MAR-1999; 99US-0123548;  
PR 23-MAR-1999; 99US-0125788;  
PR 25-MAR-1999; 99US-0126264;  
PR 29-MAR-1999; 99US-0126785;  
PR 01-APR-1999; 99US-0127462;  
PR 06-APR-1999; 99US-0128234;  
PR 08-APR-1999; 99US-0128714;  
PR 16-APR-1999; 99US-0129845;  
PR 19-APR-1999; 99US-0130077;  
PR 21-APR-1999; 99US-0130449;  
PR 23-APR-1999; 99US-0130510;  
PR 23-APR-1999; 99US-0130891;  
PR 28-APR-1999; 99US-0131449;  
PR 30-APR-1999; 99US-0132048;  
PR 30-APR-1999; 99US-0132407;  
PR 04-MAY-1999; 99US-0132484;  
PR 05-MAY-1999; 99US-0132485;

PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138647.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145219.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.

PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
Query Match		
Best Local Similarity	35.3%;	Score 946; DB 21;
Matches 198; Conservative	41.8%; Pred. No. 2,1e+75;	Length 444;
	77; Mismatches 161;	Indels 38; Gaps 77;

	Query Match	35.3%	Score 946:	DB 21:	Length 444:
	Best Local Similarity	41.8%	Pred. No.2.1e-75:		
	Matches 198:	Conservative 77:	Mismatches 161:	Indels 38:	Gaps
Qy	39	LPRKSGMPELGETIGYIKRYATATLGDFFMOOHVSKYKGIYRSNLFGEPRTIVSADAGLNR	98		
Db	3	lprpsgjlgrlletftqjlgayktpenpfrideravarysvfmhllfgeprriffaadrpetnr	62		
Qy	99	FIIDNEGELFPCSPRSTIGIGLGMKSMYLVLGDMHROMRISLINFSLHALLRTILDKDVE	158		
Db	63	fvldnegkrlfescypasiscnljgkhslllmkgsjrhkxhmbdltnsfansslkdhldmid	122		
Qy	159	RHTLFVLDWMOONSIFSAODEAKFFKFFNLAKHINMSMDPEEETEQAKKEUYFPMKGVVS	218		
Db	123	rlvfnldswesryvl--meekakitfelcvkqdmstfdrg-ewsesrlkcyllviegffs	179		
Qy	219	APRLNPLGAIYHAKALQASRTILKFERKMERKADIKEDDEEBEYKTEDAEKSKSDHNR	278		
Db	180	lrlprfstcytkarqar-----rkvaaellyvmnkkrteeaeage-----	219		
Qy	279	KQRTDDDLGVLWHLNSLSTEQILDLILSLFAGNETHSYVALIAITFLQACRAVEELR	338		
Db	220	--rkdkmaaalaaadddgfscdeavdflvalllvaggteftlmrlavkflcptrlalaqk	277		
Qy	339	EEHLEIARAKKELESESLNMDYDKKMPQOCVINEPLRLCNVYRFLHRAKLKQVRKGYD	398		
Db	278	eeheklrtamksd--syslswsdyfsmptfcgvuelrtailvaalllgqvtrtamtvelakgk	335		
Qy	399	IPSGKVLPLVISAHVHLNLSRSDQFNILPNPMMQOONNGASSGGSFSTGNN-YMPFG	457		
Db	336	lpkgwkvfsftravhldpnhldatfnpwrq-----snstvtrpsnvfprfg	385		
Qy	458	GPRCLASGELAKLEMAVFINHLVLKFWMLAEEDDQRPAPFVDFPNCGLPIRVSR	511		
Db	386	prclpogylavalaavfllhvlvgfsvsaaeqdklvftrftrtkqkryplfrivr	439		

RESULT	8
AA645023	
ID	AA645023 standard; Protein: 444 AA.
XX	
AC	AA645023;
XX	
DT	18-OCT-2000 (first entry)
XX	
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 56470.
XX	
KW	Protein identification; signal transduction pathway; metabolic pathway; hydridisation assay; genetic mapping; gene expression control; promoter termination sequence.
KW	
XX	
CS	Arabidopsis thaliana.
XX	
PM	EP1033405-A2.
XX	
PD	06-SEP-2000.
XX	

XX	25-FEB-2000	2000EED-0301439	
PR	25-FEB-1999	9905-0121825	
PR	05-MAR-1999	9905-0123180	
PR	03-MAR-1999	9905-0123548	
PR	23-MAR-1999	9905-0125788	
PR	25-MAR-1999	9905-0125654	
PR	29-MAR-1999	9905-0126785	
PR	01-APR-1999	9905-0127462	
PR	06-APR-1999	9905-0128234	
PR	16-APR-1999	9905-0128714	
PR	18-APR-1999	9905-0129845	
PR	21-APR-1999	9905-0130077	
PR	23-APR-1999	9905-0130449	
PR	23-APR-1999	9905-0130510	
PR	28-APR-1999	9905-0131449	
PR	30-APR-1999	9905-0132048	
PR	30-APR-1999	9905-0132407	
PR	04-MAY-1999	9905-0132484	
PR	05-MAY-1999	9905-0132485	
PR	06-MAY-1999	9905-0132487	
PR	06-MAY-1999	9905-0132486	
PR	07-MAY-1999	9905-0132863	
PR	11-MAY-1999	9905-0134256	
PR	14-MAY-1999	9905-0134218	
PR	14-MAY-1999	9905-0134219	
PR	14-MAY-1999	9905-0134221	
PR	14-MAY-1999	9905-0134270	
PR	16-MAY-1999	9905-0134768	
PR	15-MAY-1999	9905-0134341	
PR	20-MAY-1999	9905-0135124	
PR	21-MAY-1999	9905-0135124	
PR	24-MAY-1999	9905-0135629	
PR	25-MAY-1999	9905-0136021	
PR	27-MAY-1999	9905-0136382	
PR	28-MAY-1999	9905-0137222	
PR	01-JUN-1999	9905-0137222	
PR	03-JUN-1999	9905-0137528	
PR	04-JUN-1999	9905-0137502	
PR	07-JUN-1999	9905-0137724	
PR	08-JUN-1999	9905-0138094	
PR	10-JUN-1999	9905-0138540	
PR	10-JUN-1999	9905-0138847	
PR	14-JUN-1999	9905-0139119	
PR	16-JUN-1999	9905-0139452	
PR	16-JUN-1999	9905-0139453	
PR	17-JUN-1999	9905-0139454	
PR	18-JUN-1999	9905-0139452	
PR	18-JUN-1999	9905-0139456	
PR	18-JUN-1999	9905-0139455	
PR	18-JUN-1999	9905-0139461	
PR	18-JUN-1999	9905-0139462	
PR	18-JUN-1999	9905-0139463	
PR	18-JUN-1999	9905-0139750	
PR	18-JUN-1999	9905-0139817	
PR	21-JUN-1999	9905-0139871	
PR	22-JUN-1999	9905-0139869	
PR	23-JUN-1999	9905-0140353	
PR	23-JUN-1999	9905-0140354	
PR	24-JUN-1999	9905-0140595	
PR	24-JUN-1999	9905-0140623	
PR	29-JUN-1999	9905-0140591	
PR	30-JUN-1999	9905-0141287	
PR	01-JUL-1999	9905-0141284	
PR	01-JUL-1999	9905-0142154	
PR	02-JUL-1999	9905-0142055	
PR	06-JUL-1999	9905-0142390	
PR	08-JUL-1999	9905-0142803	





```
RESULT      9
AAG46490
ID    AAG46490 standard; Protein; 461 AA.
XX
AC    AAG46490;
XX
DT    18-OCT-2000 (first entry)
XX
DE    Arabidopsis thaliana protein fragment SEQ ID NO: 58494.
XX
KW    Protein identification; signal transduction pathway; metabolic pathway;
KW    hybridisation assay; genetic mapping; gene expression control; promoter;
KW    termination sequence.
XX
OS    Arabidopsis thaliana.
XX
PN    EP1033405-A2.
XX
PD    06-SEP-2000.
XX
PE    25-FEB-2000; 2000EP-0301439.
XX
PR    25-FEB-1999; 99US-0121825.
PR    05-MAR-1999; 99US-0123180.
PR    09-MAR-1999; 99US-0123548.
PR    23-MAR-1999; 99US-0125788.
PR    23-MAR-1999; 99US-0126264.
PR    29-MAR-1999; 99US-0126785.
PR    01-APR-1999; 99US-0127462.
PR    06-APR-1999; 99US-0128234.
PR    08-APR-1999; 99US-0128714.
PR    16-APR-1999; 99US-0129845.
PR    19-APR-1999; 99US-0130077.
PR    21-APR-1999; 99US-0130449.
PR    23-APR-1999; 99US-0130510.
PR    28-APR-1999; 99US-0130891.
PR    30-APR-1999; 99US-0132048.
PR    30-APR-1999; 99US-0132407.
PR    04-MAY-1999; 99US-0132484.
PR    05-MAY-1999; 99US-0132485.
PR    06-MAY-1999; 99US-0132486.
PR    06-MAY-1999; 99US-0132487.
PR    11-MAY-1999; 99US-0132863.
PR    11-MAY-1999; 99US-0134256.
PR    14-MAY-1999; 99US-0134218.
PR    14-MAY-1999; 99US-0134219.
PR    14-MAY-1999; 99US-0134221.
PR    14-MAY-1999; 99US-0134370.
PR    18-MAY-1999; 99US-0134768.
PR    19-MAY-1999; 99US-0134941.
PR    20-MAY-1999; 99US-0135124.
PR    21-MAY-1999; 99US-0135353.
PR    24-MAY-1999; 99US-0135629.
PR    25-MAY-1999; 99US-0136021.
PR    27-MAY-1999; 99US-0136382.
PR    28-MAY-1999; 99US-0136782.
PR    01-JUN-1999; 99US-0137222.
PR    03-JUN-1999; 99US-0137528.
PR    04-JUN-1999; 99US-0137502.
PR    07-JUN-1999; 99US-0137724.
PR    08-JUN-1999; 99US-0138094.
PR    10-JUN-1999; 99US-0138540.
PR    10-JUN-1999; 99US-0138847.
PR    14-JUN-1999; 99US-0139119.
PR    16-JUN-1999; 99US-0139452.
PR    16-JUN-1999; 99US-0139453.
PR    17-JUN-1999; 99US-0139482.
PR    18-JUN-1999; 99US-0139454.
PR    18-JUN-1999; 99US-0139455.
PR    18-JUN-1999; 99US-0139456.
PR    18-JUN-1999; 99US-0139457.
PR    18-JUN-1999; 99US-0139458.
PR    18-JUN-1999; 99US-0139459.
PR    18-JUN-1999; 99US-0139460.
PR    18-JUN-1999; 99US-0139461.
PR    18-JUN-1999; 99US-0139462.
PR    18-JUN-1999; 99US-0139463.
PR    18-JUN-1999; 99US-0139750.
PR    18-JUN-1999; 99US-0139763.
PR    21-JUN-1999; 99US-0139817.
PR    22-JUN-1999; 99US-0139899.
PR    23-JUN-1999; 99US-0140353.
PR    23-JUN-1999; 99US-0140354.
PR    24-JUN-1999; 99US-0140695.
PR    28-JUN-1999; 99US-0140823.
PR    29-JUN-1999; 99US-0140991.
PR    30-JUN-1999; 99US-0141287.
PR    01-JUL-1999; 99US-0141842.
PR    01-JUL-1999; 99US-0142154.
PR    02-JUL-1999; 99US-0142055.
PR    06-JUL-1999; 99US-0142380.
PR    08-JUL-1999; 99US-0142803.
PR    09-JUL-1999; 99US-0142920.
PR    12-JUL-1999; 99US-0142977.
PR    13-JUL-1999; 99US-0143542.
PR    14-JUL-1999; 99US-0143624.
PR    15-JUL-1999; 99US-0144005.
PR    16-JUL-1999; 99US-0144085.
PR    16-JUL-1999; 99US-0144086.
PR    19-JUL-1999; 99US-0144325.
PR    19-JUL-1999; 99US-0144332.
PR    19-JUL-1999; 99US-0144333.
PR    19-JUL-1999; 99US-0144334.
PR    19-JUL-1999; 99US-0144335.
PR    20-JUL-1999; 99US-0144352.
PR    20-JUL-1999; 99US-0144632.
PR    20-JUL-1999; 99US-0144684.
PR    21-JUL-1999; 99US-0144814.
PR    21-JUL-1999; 99US-0145086.
PR    21-JUL-1999; 99US-0145088.
PR    22-JUL-1999; 99US-0145087.
PR    22-JUL-1999; 99US-0145088.
PR    22-JUL-1999; 99US-0145089.
PR    22-JUL-1999; 99US-0145192.
PR    23-JUL-1999; 99US-0145145.
PR    23-JUL-1999; 99US-0145218.
PR    23-JUL-1999; 99US-0145224.
PR    26-JUL-1999; 99US-0145276.
PR    27-JUL-1999; 99US-0145913.
PR    27-JUL-1999; 99US-0145918.
PR    27-JUL-1999; 99US-0145919.
PR    28-JUL-1999; 99US-0145951.
PR    02-AUG-1999; 99US-0146386.
PR    02-AUG-1999; 99US-0146388.
PR    02-AUG-1999; 99US-0146389.
PR    03-AUG-1999; 99US-0147038.
PR    04-AUG-1999; 99US-0147204.
PR    04-AUG-1999; 99US-0147302.
PR    05-AUG-1999; 99US-0147309.
PR    05-AUG-1999; 99US-0147260.
PR    05-AUG-1999; 99US-0147303.
PR    06-AUG-1999; 99US-0147416.
PR    09-AUG-1999; 99US-0147493.
PR    09-AUG-1999; 99US-0147935.
PR    10-AUG-1999; 99US-0148171.
PR    11-AUG-1999; 99US-0148319.
PR    12-AUG-1999; 99US-0148341.
PR    13-AUG-1999; 99US-0148565.
PR    13-AUG-1999; 99US-0148684.
PR    16-AUG-1999; 99US-0149368.
PR    17-AUG-1999; 99US-0149175.
PR    18-AUG-1999; 99US-0149426.
PR    20-AUG-1999; 99US-0149722.
```

PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	13-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161400.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

	Query Match	27.1%	Score 726;	DB 21;	Length 461; .
	Best Local Similarity	32.3%	Pred. No.	6.6e-56;	
	Matches 166;	Conservative	91;	Mismatches 181;	Indels 76; Gaps 10;
Oy	12 LLLPSSLSLLEFLILKKRN--RKTRFNPGRSGMPLEGITGYLKPKYATTLGDGFMQ	69			
Dd	7 ILLISLCgall-----rwgmqrlytknglppptmgwplfgetetflkygp-----nfrr	55			
Oy	70 QHVSRYKIRTSNLTGEPTIVSADAGLNRFILQNGLRECSYPSRIGIQLKGSMELVLY	129			
Dd	56 nqrlrysgffkehlhgpcrllsmdeevnryllkneskgvlpypgpmldilgtctmaavh	115			
Oy	130 GDMDHDMSISINFSHARLRITLIKDYERHTLFVLDSMQONSITSQAQDEAKRTFF---	185			
Dd	116 gsshtlmrgslslissumtchtlipkvdhmrsvldgwnelevididktthmalfssl	175			

QY	166	----	NLAAKHTMSMDPEEETEOQKKYVFMFGVVSAPRLNLPCTVYHKALOSRAITLK	240
DB	176	lqiaqlrkrfv-----	eeftkafkflvvgclsvpidlpgltnyrcsqiqatnmidr	225
QY	241	FIERMEERKLDIKEDDEDEEEVKTDEDAEMSKSDHWKQRTDDDLGAWLKHSN----	L 296	
DB	226	llreImger-----	dage-----tfcdmlylmkkggnrypl	258
QY	297	STEQIDLIDLLIFAGHETSSVAIALALFLOACPKAVEELREHLEIARAKKELGSESL	356	
DB	259	tdeairdgvvtcltysgytsvstsmmaiklylhdhpkalqelraehlafrerkiq--depl	316	
QY	357	NMDDVKKMDFNQCVENELRLGNVYRFLRKALKDVRKKGVDPSGKVLRYLSAVHLDN	416	
DB	317	gledcksmkfravlyetstrlatlvngvlrkttrdldngyllpkrwrlvyutrelngda	376	
QY	417	SRVDDPFLNFPMRWQOONNGASSSGSGSFSTWGNMYPFGGPPRLCAGSELAKLEMAVFI	476	
DB	377	nlyeprllfnprwmkksleeg-----	nsctfvf9ggtclcpkgkelgivelssfl	425
QY	477	HHLVLEKFWELAEDDQFPAPFPVDEPNGLPIRVAS	510	
DB	426	hyfvttryweeigdelmvfrvtfapkgfhlris	459	
RESULT	10			
AAAG46489				
ID	AAAG46489	standard; Protein; 462 AA.		
XX	AAAG46489;			
DT	18-OCT-2000	(first entry)		
XX	Arabidopsis thaliana protein fragment SEQ ID NO: 58493.			
KW	Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter termination sequence.			
OS	Arabidopsis thaliana.			
XX	EpI033405-A2.			
PN	EpI033405-A2.			
PD	06-SEP-2000.			
XX				
PF	25-FEB-2000; 2000EP-0301439.			
XX				
PR	25-FEB-1999;	99US-0121825.		
PR	05-MAR-1999;	99US-0123180.		
PR	09-MAR-1999;	99US-0133548.		
PR	23-MAR-1999;	99US-0125788.		
PR	25-MAR-1999;	99US-0126284.		
PR	29-MAR-1999;	99US-0126785.		
PR	01-APR-1999;	99US-0127462.		
PR	06-APR-1999;	99US-0128234.		
PR	08-APR-1999;	99US-0128714.		
PR	16-APR-1999;	99US-0129845.		
PR	19-APR-1999;	99US-0130077.		
PR	21-APR-1999;	99US-0130449.		
PR	23-APR-1999;	99US-0130510.		
PR	23-APR-1999;	99US-0130891.		
PR	28-APR-1999;	99US-0131449.		
PR	30-APR-1999;	99US-0132048.		
PR	04-MAY-1999;	99US-0132407.		
PR	05-MAY-1999;	99US-0132484.		
PR	06-MAY-1999;	99US-0132485.		
PR	06-MAY-1999;	99US-0132486.		
PR	07-MAY-1999;	99US-0132487.		
PR	11-MAY-1999;	99US-0132853.		
PR	14-MAY-1999;	99US-0134256.		
PR	14-MAY-1999;	99US-0134218.		
PR	14-MAY-1999;	99US-0134219.		

PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.

PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149920.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 18-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.

PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 27.18; Score 726; DB 21; Length 462;  
Best Local Similarity 32.38; Pred. No. 6,6e-56;

Matches 166; Conservative 91; Mismatches 181; Indels 76; Gaps 10;

QY 12 LLLPSLLSLFLFLRLRN--RKTRFNLPKSGKGFLEETGYLKYATATLIGDEM 69  
DB 8 LLLVLSLsall-----rwmgytknglppgtmgwplfgeteifkqgp-----nfmr 56  
QY 70 QHVSRYKIVYSNLFEGEPTIVSADAGLNRFLLQNEGRLECSYPRISGILGKSMVLV 129  
DB 57 ngrlrygffkshllgpcplismdsevnryllkneskglvpgyqsmldlgtcmaavh 116  
QY 130 GDMHRDMSISLNFLSHARKLRTILKDYERHTLPVLDWMQNSIFSADEAKKTF---- 185  
DB 117 gssnrlmgsllstlssctmmrdhllpkvdhmrsvldqwnelavldqdkckhmaflssl 176  
QY 186 -----NLMAKIMSMDPGEETEOLKREYTEMKGVASAPLNLPQTAVHAKALOSRATILK 240  
DB 177 tqiagnllkprty-----eeftakfklvgtlsvpldipgtnyrgtqatnmidr 226  
QY 241 FIERKMEERKLDIEEDQEEVEEVTEDAEKMSKSDHVRKORTDDLLGWLKHSN---L 296  
DB 227 llreimgerr-----dege-----tfdmlylmlkkggnrypl 259  
QY 297 STEQILDILSLFRAGHTSSVATAALAFFLQACPKAVELEREHLEFLARKKELGESEL 356  
DB 260 ldeetrdgvlllysgyeltstsmaklyldhpkalqelraefrekrq--depl 317  
QY 357 NMDYKKKMDFTQCVINETLRGNVVRFLHRKALDKVRYKGYDIPSGWKVLPVISAVIDN 416  
DB 318 gledkxsmkfravlyetsrlatlvngylrtctrdleingylipkgyviyyteingda 377  
QY 417 SRYDQPNLFNPRMOQONNGASSSGSSTWGNVMPFGGPGPRLCAGSELAKLEMAVEI 476  
DB 378 nlyceplrlfnpwrwmkksleeg-----nscfvgggtlrlpckelgelveisfl 426  
QY 477 HHLVKEWELAEDDQPAFPFVDFPNGLPIRVS 510  
DB 427 hyflvryrweelggdelmvrprvfapkgfhlrls 460

## RESULT 11

AAG46491  
ID AAG46491 standard; Protein: 465 AA.

AC AAG46491;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 58495.

KW Protein identification; signal transduction pathway; metabolic pathway;  
hybridisation assay; genetic mapping; gene expression control; promoter;  
termination sequence.

OS Arabidopsis thaliana.

PN EPI033405-A2.

PD 06-SEP-2000.

PE 25-FEB-2000; 2000EP-0301439.

XX

PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139753.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 26-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142380.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 26-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.

PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 27.1%; Score 726; DB 21; Length 465;  
Best Local Similarity 32.3%; Pred. No. 6; 76-56;  
Matches 166; Conservative 91; Mismatches 181; Indels 76; Gaps 10;

QY 12 LLLPSSLILFLILKRRN--RKRPNLPPGSGMPLEGTGYLKPATYTLGDMQ 69  
DB 11 LLLPSSLILFLILKRRN--RKRPNLPPGSGMPLEGTGYLKPATYTLGDMQ 59  
QY 70 QHVSXKIVRNSLFGERTIYSADAGLNRTLLONGRLEFCSYPSISGILGKMSMLV 129  
DB 60 ngrllygsffkshllygcplismdsevrlylkneskylvpypqsmldilgctmmaavh 119  
QY 130 GDMHRMRSISLNFSLHARLRTILKDYERHTLFLVLDGMOONSIFSADEAKKFTF 185  
DB 120 gsshrlnrsgllslstssummrchllpkvdhmrsgldqwnelavldidcktkhmaflsl 179  
QY 186 ----NLAKHIMSMDPGEETEDLKKYVTFPMGVASAPLNLPSTAYNAKALQSRATLK 240  
DB 180 tqiagnlrfkprv-----eeftktafklyvgclsvrldlprtnyrgsgiqarnidr 229  
QY 241 FIERKMEERKIDIKREDEDEEEVKTEDAEKMSKSHVAKKORTDOLGLMVLNKHSN 296  
DB 230 llrelmgerr-----dsge-----tftmdlylmkkggnrpyl 262  
QY 297 STEQILDLILSLFAGHETSSVAIALAFLOACPKAVEELREHLEIETARAKKEGESEL 356  
DB 263 tdeairdgvtrlllysgyvetstsmalkylhdpkaiqlraehlafrckrg--depl 320  
QY 357 NMDYKKMDFTOCVINFTLRIGNVVFLHRAKALDVRKKGIDIPSGWLVLPVISAVIDN 416  
DB 321 gldvksmkftravlyetsrlatinyvlrktctdlnsgylipgyrilyytrelnyda 380  
QY 417 SRYDOPNLFNPMWQOONNGASSSGSFSITWGNVMPFGGPRCAGSELAKLEMAVEI 476  
DB 381 nlyedpdlifnprwmkkslesq-----nscftvgggtlcpkqelgiveisfl 429  
QY 477 HHLVLKFNWELAEEDQPPAFPFVDFPNCIPRVS 510  
DB 429 HHLVLKFNWELAEEDQPPAFPFVDFPNCIPRVS 510

Db 430 hYfVrYrweei9gdelmvfrvFapKqfhlris 463  
RESULT 12  
ID AAG11836 standard; Protein: 461 AA.  
XX AAG11836;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 10708.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 16-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 26-JUL-1999; 99US-0145224.  
PR 27-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.

```

PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0156599.
PR 29-SEP-1999; 99US-0156458.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 26.6%; Score 712; DB 21; Length 461;  
 Best Local Similarity 32.5%; Pred. No. 1,le-54;  
 Matches 164; Conservative 93; Mismatches 196; Indels 52; Gaps 8;

```

QY 12 LLLPSSLLELFLILKRRN--KRTRENLPFGKSGPFLGCTTYLKYPTATTLGDFMQ 69
DB 1 mmllgllylvcltallrwmqmyskglppgtmgwplfgeteflkgp-----dfmk 55
QY 70 QHVSYKGIYSNLCGEPTIYADAGLRFILQNGRLEFECSPYSIGIGLKGWMLVY 129
DB 56 ngrlrygsffshlqgcpltymdelnryllmeskglvagypgsmldllytclnaaah 115
QY 130 GDMHRDMKSISLNFSLHARLFTILKDVERTLFLVLDWMQONSIPSADEAKKFTFNMA 189

```

```

DB 116 gpshtlmrgslslslspsrmkdhllpkiddfmrnlylgcwddletvdqekthmafi-1ss 174
QY 190 KHMSMDPEEETEOJLKEVYTFMGVVSAPLNLPGYVHKAQSRATILKEKMEER 249
DB 175 llqlaelklkpeveeyreteflkvglvtlsvpdipgtnyrsyqvarmndrilltelmqer 234
QY 250 KLDIKEEDQEEVEEVTEDEAEKSKSDHYRKQRTDDDLGWTAKHN-----LSTEQILDLI 305
DB 235 k-----esgetft-----dmlylmkkednrylltdkeirgqv 267
QY 306 LSLPAGHETSSVAIALAIFLQACPKAVEELREBHLELARAKKKLGESELMWDDYKKMD 365
DB 268 vtllysgyetsvtcmmalkylhdpkpkaleelrrehlairekrp--depltdldikemk 325
QY 366 FTQCVINETLRGNVRFPLHRRKALKDVRKGYDIPSGMKVLPVISAHLNDSRYDQPLF 425
DB 326 flravlfetsrlatlvngvlrktthdlnyglipkgywlyyutrelngdtslyedpmif 385
QY 426 NPMRWQOQNNNGASSSGSFSFTWGNMYMPFGGPRLLCGSELAKLEMAVFTHILKFNW 485
DB 386 npwrmekslesks-----yflfg9vrlcpqkelgisevsflnyftvkyrw 434
QY 486 ELAEDDQFAFPFVDFPGLPIRVS 510
DB 435 geengedklmfvprsapkgyhlcxs 459

RESULT 13
AAG11835
ID AAG11835 standard; Protein; 462 AA.
XX
AC AAG11835;
DE 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 10707.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
OS
XX
XX EPI033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 04-MAY-1999; 99US-0132407.
XX 05-MAY-1999; 99US-0132484.
XX 06-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 07-MAY-1999; 99US-0132487.
XX 11-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.

```

PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 14-MAY-1999; 99US-0134376.  
PR 15-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140981.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.

PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0156559.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159684.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.



PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
PR

Query Match	26.6%;	Score 712;	DB 21;	Length 462;
Post Local Similarity	33.5%;	Prod No 1	10-FA	

Matches 164; Conservative 93; Mismatches 196; Indels 52; Gaps 8;

```

OY 12 LLLPPLSLSLFLLKLRN - RKTFRNLPCKSGMPFGETIGYKLPATANTLQPMQ 69
Db 2 mmlgllylvccltallrwngmtysskglpbgmpmglfgettefikgp-----dfmk 56
OY 70 QHVSRYGKIYRNLGEBPTIVSADAGLNRFLLONEGRLEFCSYPRISIGGILGKWSMLV 129
Db 57 nqrlrysifkshllpocltvsmdaelnrylimeskgilvaypqsmlldlqcnlaavh 116
OY 130 GDMHRDMSISLNFSLHARLRTLLTKOVERTLTVLBSWOONSJFSNODAKKFTFLMA 189
Db 117 gprhrlmgslslslstpmkdhllpkldfmrnylogwdletvdikextklnaf-iss 175
OY 190 KHMSMDPGEETBOLKKEYVTFKGVVSAPLNPGTAHYKALOSRATILKFTERKEER 249
Db 176 llqiaetlkrpeveeytreffkllvgvclsvldipglnfysgvqarnidrlldtelmger 235
OY 250 KLIDKEQDEEVEKTEDEAMSKMSDHVRKORODDLGNWLKHSN---LSTFOIIDL 305
Db 236 k-----esgeftf-----omdylmkbednrylldelrldqv 268
OY 306 LSLFAGHETSSVALAIFLIACPRABEELREHEELIRAKKELGESLEIMBDDYKMD 365
Db 269 vtllysgyeltvstsmalklphdhpakleeltrahlatrerkrp--depltdlksmk 326
OY 366 FTQCVINETLALGNVRFRLHAKKALKOVRYKGYDIPSGWKVLPVTSAYHLNDSRKIDQNL 423
Db 337 ftrvrlfetsrlatlvngvrlktthldiengylllpkwrlyvvytrelnyhtslsyedpmif 366
OY 426 NPMWQOONNASSSGSGSEFTWNNMMPGEGRCGLRAGSELAKLEAVALTHNLVLFKN 485
Db 387 npwrmekslesks-----yflfsgyvrldpkyelgsevsfhlhyvtekyzw 435
OY 486 ELAEDDQFAFPFVDFPGLPIRVS 510
Db 436 eengedklmvrpvsapkyhllks 460

```

RESULT	14
AA01834	
ID	AA01834 standard; Protein; 465 AA

AC AAG11834;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 10706

KM Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.

OS *Arabidopsis thaliana*.

PN EP1033405-A2.

PD 06-SEP-2000.

XX

PR	25-FEB-2000;	2000EP-0301433.	
PR	25-FEB-1999;	9905-0121825.	
PR	05-MAR-1999;	9905-0123180.	
PR	09-MAR-1999;	9905-0123548.	
PR	23-MAR-1999;	9905-0125788.	
PR	25-MAR-1999;	9905-0126264.	
PR	29-MAR-1999;	9905-0126785.	
PR	01-APR-1999;	9905-0127462.	
PR	06-APR-1999;	9905-0128234.	
PR	16-APR-1999;	9905-0128714.	
PR	19-APR-1999;	9905-0128845.	
PR	21-APR-1999;	9905-0130077.	
PR	23-APR-1999;	9905-0130449.	
PR	28-APR-1999;	9905-0130510.	
PR	30-APR-1999;	9905-0130891.	
PR	03-MAY-1999;	9905-0131449.	
PR	04-MAY-1999;	9905-0132048.	
PR	05-MAY-1999;	9905-0132407.	
PR	06-MAY-1999;	9905-0134884.	
PR	07-MAY-1999;	9905-0134851.	
PR	11-MAY-1999;	9905-0134866.	
PR	14-MAY-1999;	9905-0134873.	
PR	14-MAY-1999;	9905-0134885.	
PR	18-MAY-1999;	9905-0134921.	
PR	19-MAY-1999;	9905-0134941.	
PR	20-MAY-1999;	9905-0135124.	
PR	21-MAY-1999;	9905-0135129.	
PR	24-MAY-1999;	9905-0135353.	
PR	25-MAY-1999;	9905-0135629.	
PR	27-MAY-1999;	9905-0136021.	
PR	28-MAY-1999;	9905-0136392.	
PR	03-JUN-1999;	9905-0136782.	
PR	04-JUN-1999;	9905-0137222.	
PR	07-JUN-1999;	9905-0137528.	
PR	04-JUN-1999;	9905-0137502.	
PR	10-JUN-1999;	9905-0137724.	
PR	16-JUN-1999;	9905-0138094.	
PR	17-JUN-1999;	9905-0138540.	
PR	18-JUN-1999;	9905-0138847.	
PR	18-JUN-1999;	9905-0139119.	
PR	18-JUN-1999;	9905-0139452.	
PR	18-JUN-1999;	9905-0139456.	
PR	18-JUN-1999;	9905-0139461.	
PR	18-JUN-1999;	9905-0139463.	
PR	18-JUN-1999;	9905-0139465.	
PR	18-JUN-1999;	9905-0139467.	
PR	18-JUN-1999;	9905-0139468.	
PR	18-JUN-1999;	9905-0139469.	
PR	18-JUN-1999;	9905-0139470.	
PR	18-JUN-1999;	9905-0139471.	
PR	18-JUN-1999;	9905-0139472.	
PR	18-JUN-1999;	9905-0139473.	
PR	18-JUN-1999;	9905-0139474.	
PR	18-JUN-1999;	9905-0139475.	
PR	18-JUN-1999;	9905-0139476.	
PR	18-JUN-1999;	9905-0139477.	
PR	18-JUN-1999;	9905-0139478.	
PR	18-JUN-1999;	9905-0139479.	
PR	18-JUN-1999;	9905-0139480.	
PR	18-JUN-1999;	9905-0139481.	
PR	18-JUN-1999;	9905-0139482.	
PR	18-JUN-1999;	9905-0139483.	
PR	18-JUN-1999;	9905-0139484.	
PR	18-JUN-1999;	9905-0139485.	
PR	18-JUN-1999;	9905-0139486.	
PR	18-JUN-1999;	9905-0139487.	
PR	18-JUN-1999;	9905-0139488.	
PR	18-JUN-1999;	9905-0139489.	
PR	18-JUN-1999;	9905-0139490.	
PR	18-JUN-1999;	9905-0139491.	
PR	18-JUN-1999;	9905-0139492.	
PR	18-JUN-1999;	9905-0139493.	
PR	18-JUN-1999;	9905-0139494.	
PR	18-JUN-1999;	9905-0139495.	
PR	18-JUN-1999;	9905-0139496.	
PR	18-JUN-1999;	9905-0139497.	
PR	18-JUN-1999;	9905-0139498.	
PR	18-JUN-1999;	9905-0139499.	
PR	18-JUN-1999;	9905-0139500.	
PR	18-JUN-1999;	9905-0139501.	
PR	18-JUN-1999;	9905-0139502.	
PR	18-JUN-1999;	9905-0139503.	
PR	18-JUN-1999;	9905-0139504.	
PR	18-JUN-1999;	9905-0139505.	
PR	18-JUN-1999;	9905-0139506.	
PR	18-JUN-1999;	9905-0139507.	
PR	18-JUN-1999;	9905-0139508.	
PR	18-JUN-1999;	9905-0139509.	
PR	18-JUN-1999;	9905-0139510.	
PR	18-JUN-1999;	9905-0139511.	
PR	18-JUN-1999;	9905-013951	

PR	28-SEP-1999;	9905-0156458.
PR	29-SEP-1999;	9905-0156596.
PR	04-OCT-1999;	9905-0157117.
PR	05-OCT-1999;	9905-0157753.
PR	06-OCT-1999;	9905-0157865.
PR	07-OCT-1999;	9905-0158029.
PR	08-OCT-1999;	9905-0158232.
PR	12-OCT-1999;	9905-0158369.
PR	13-OCT-1999;	9905-0159293.
PR	13-OCT-1999;	9905-0159294.
PR	13-OCT-1999;	9905-0159295.
PR	14-OCT-1999;	9905-0159329.
PR	14-OCT-1999;	9905-0159330.
PR	14-OCT-1999;	9905-0159331.
PR	14-OCT-1999;	9905-0159637.
PR	14-OCT-1999;	9905-0159638.
PR	18-OCT-1999;	9905-0159584.
PR	21-OCT-1999;	9905-0160741.
PR	21-OCT-1999;	9905-0160767.
PR	21-OCT-1999;	9905-0160768.
PR	21-OCT-1999;	9905-0160770.
PR	21-OCT-1999;	9905-0160814.
PR	21-OCT-1999;	9905-0160815.
PR	22-OCT-1999;	9905-0160980.
PR	22-OCT-1999;	9905-0160981.
PR	22-OCT-1999;	9905-0160989.
PR	25-OCT-1999;	9905-0161404.
PR	25-OCT-1999;	9905-0161405.
PR	25-OCT-1999;	9905-0161406.
PR	26-OCT-1999;	9905-0161359.
PR	26-OCT-1999;	9905-0161360.
PR	26-OCT-1999;	9905-0161361.
PR	28-OCT-1999;	9905-0161920.
PR	28-OCT-1999;	9905-0161922.
PR	28-OCT-1999;	9905-0161993.
PR	29-OCT-1999;	9905-0162142.

Query Match	26.68;	Score 712;	DB 21;	Length 465;
Best Local Similarity	32.18;	Pred. NO. 1.2e-54;		
Matches 163; Conservative	94;	Mismatches 194;	Indels 56;	Gaps 8;

[illegible]

OY 484 NWELAEEDDPFAFFVDFPNGLPPIRV 510  
Db 437 rweengedkilmvfrvsapsyhlkcs 463

## RESULT 15

AAG20783

AAG20783 standard; Protein: 468 AA.

AAG20783; 17-OCT-2000 (first entry)

Arabidopsis thaliana protein fragment SEQ ID NO: 23102.

Protein identification: signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.

Arabidopsis thaliana.

EP1033405-A2.

25-FEB-2000; 2000EP-0301439.

25-FEB-1999; 99US-0121825.

05-MAR-1999; 99US-0123180.

09-MAR-1999; 99US-0123548.

23-MAR-1999; 99US-0125788.

29-MAR-1999; 99US-0126284.

01-APR-1999; 99US-0126785.

06-APR-1999; 99US-0127462.

16-APR-1999; 99US-0128714.

21-APR-1999; 99US-0129845.

23-APR-1999; 99US-0130077.

23-APR-1999; 99US-0130449.

23-APR-1999; 99US-0130510.

28-APR-1999; 99US-0130891.

30-APR-1999; 99US-0131449.

04-MAY-1999; 99US-0132048.

05-MAY-1999; 99US-0132484.

06-MAY-1999; 99US-0132485.

07-MAY-1999; 99US-0132486.

11-MAY-1999; 99US-0132863.

14-MAY-1999; 99US-0134218.

14-MAY-1999; 99US-0134219.

14-MAY-1999; 99US-0134221.

18-MAY-1999; 99US-0134370.

19-MAY-1999; 99US-0134768.

20-MAY-1999; 99US-0134941.

21-MAY-1999; 99US-0135124.

24-MAY-1999; 99US-0135353.

25-MAY-1999; 99US-0136021.

27-MAY-1999; 99US-0136392.

28-MAY-1999; 99US-0136782.

03-JUN-1999; 99US-0137222.

04-JUN-1999; 99US-0137528.

07-JUN-1999; 99US-0137724.

08-JUN-1999; 99US-0138094.

10-JUN-1999; 99US-0138540.

10-JUN-1999; 99US-0138847.

14-JUN-1999; 99US-0139119.

16-JUN-1999; 99US-0139452.

17-JUN-1999; 99US-0139453.

17-JUN-1999; 99US-0139492.

18-JUN-1999; 99US-0139454.

18-JUN-1999; 99US-0139455.

18-JUN-1999; 99US-0139456.

18-JUN-1999; 99US-0139457.

18-JUN-1999; 99US-0139458.

18-JUN-1999; 99US-0139459.

18-JUN-1999; 99US-0139460.

18-JUN-1999; 99US-0139461.

18-JUN-1999; 99US-0139462.

18-JUN-1999; 99US-0139463.

18-JUN-1999; 99US-0139750.

18-JUN-1999; 99US-0139763.

21-JUN-1999; 99US-0139817.

22-JUN-1999; 99US-0139899.

23-JUN-1999; 99US-0140353.

23-JUN-1999; 99US-0140354.

24-JUN-1999; 99US-0140695.

28-JUN-1999; 99US-0140823.

29-JUN-1999; 99US-0140991.

30-JUN-1999; 99US-0141287.

01-JUL-1999; 99US-0141842.

01-JUL-1999; 99US-0142154.

02-JUL-1999; 99US-0142055.

06-JUL-1999; 99US-0142390.

08-JUL-1999; 99US-0142803.

09-JUL-1999; 99US-0142920.

12-JUL-1999; 99US-0142977.

13-JUL-1999; 99US-0143542.

14-JUL-1999; 99US-0143624.

15-JUL-1999; 99US-0144005.

16-JUL-1999; 99US-0144085.

19-JUL-1999; 99US-0144086.

19-JUL-1999; 99US-0144325.

19-JUL-1999; 99US-0144331.

19-JUL-1999; 99US-0144332.

19-JUL-1999; 99US-0144333.

19-JUL-1999; 99US-0144334.

20-JUL-1999; 99US-0144335.

20-JUL-1999; 99US-0144352.

20-JUL-1999; 99US-0144632.

20-JUL-1999; 99US-0144884.

21-JUL-1999; 99US-0144814.

21-JUL-1999; 99US-0145086.

21-JUL-1999; 99US-0145088.

22-JUL-1999; 99US-0145085.

22-JUL-1999; 99US-0145087.

22-JUL-1999; 99US-0145089.

22-JUL-1999; 99US-0145192.

23-JUL-1999; 99US-0145145.

23-JUL-1999; 99US-0145218.

23-JUL-1999; 99US-0145224.

26-JUL-1999; 99US-0145276.

27-JUL-1999; 99US-0145913.

27-JUL-1999; 99US-0145918.

27-JUL-1999; 99US-0145919.

28-JUL-1999; 99US-0145919.

02-AUG-1999; 99US-0145951.

02-AUG-1999; 99US-0146386.

02-AUG-1999; 99US-0146388.

02-AUG-1999; 99US-0146389.

03-AUG-1999; 99US-0147038.

04-AUG-1999; 99US-0147204.

04-AUG-1999; 99US-0147204.

05-AUG-1999; 99US-0147302.

05-AUG-1999; 99US-0147302.

05-AUG-1999; 99US-0147320.

05-AUG-1999; 99US-0147260.

06-AUG-1999; 99US-0147303.

06-AUG-1999; 99US-0147416.

09-AUG-1999; 99US-0147493.

09-AUG-1999; 99US-0147935.

10-AUG-1999; 99US-0148171.

11-AUG-1999; 99US-0148319.

12-AUG-1999; 99US-0148341.

13-AUG-1999; 99US-0148565.

13-AUG-1999; 99US-0148684.



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: October 17, 2001, 19:19:21 ; Search time 26.13 Seconds  
(without alignments)  
404.242 Million cell updates/sec

Title: US-09-502-426-2

Perfect score: 2681  
Sequence: 1 MEETHTHLLPLLLPILLS.....FAFPVDPNGLPPIVSRLL 513

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PTOUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	978.5	36.5	472	2	US-08-622-166A-2
2	978.5	36.5	472	2	US-08-622-166A-2
3	503	18.8	492	3	US-08-724-466B-2
4	439	16.4	497	3	US-08-724-466B-4
5	289	10.8	511	4	US-08-991-677-4
6	279.5	10.4	513	3	US-08-948-564-6
7	264	9.8	520	2	US-09-091-432-2
8	263.5	9.8	504	1	US-08-457-274A-25
9	263.5	9.8	504	5	PCT-US95-05758-25
10	257	9.6	476	1	US-08-313-075A-30
11	256.5	9.6	480	1	US-08-201-118-7
12	256.5	9.6	480	2	US-08-238-821B-7
13	256.5	9.6	490	5	PCT-US95-05744-7
14	245	9.1	495	1	US-08-532-065B-2
15	244.5	9.1	576	3	US-08-948-564-16
16	241	9.0	510	3	US-08-606-505B-66
17	241	9.0	510	4	US-09-616-990-66
18	240.5	9.0	509	3	US-08-948-564-18
19	239	8.9	496	1	US-08-313-075A-50
20	236	8.8	521	3	US-08-948-564-14
21	234.5	8.7	490	1	US-08-201-118-1
22	234.5	8.7	480	2	US-08-238-821B-1
23	234.5	8.7	480	5	PCT-US95-05744-1
24	233.5	8.7	508	4	US-08-991-677-2
25	231	8.6	506	1	US-08-313-075A-38
26	231	8.6	506	3	US-08-606-505B-65
27	231	8.6	506	4	US-09-616-990-65

28	223.5	8.3	496	4	US-09-292-768-64	Sequence 64, Appl
29	223	8.3	490	1	US-08-201-118-3	Sequence 3, Appl
30	223	8.3	490	2	US-08-238-821B-3	Sequence 3, Appl
31	223	8.3	490	5	PCT-US95-05744-3	Sequence 3, Appl
32	222	8.3	490	1	US-08-201-118-9	Sequence 9, Appl
33	222	8.3	490	2	US-08-238-821B-9	Sequence 9, Appl
34	222	8.3	490	5	PCT-US95-05744-9	Sequence 9, Appl
35	222	8.3	516	3	US-08-948-564-12	Sequence 12, Appl
36	221.5	8.3	500	2	US-08-314-601-2	Sequence 2, Appl
37	221.5	8.3	500	5	PCT-US95-13051-2	Sequence 2, Appl
38	218	8.1	510	3	US-08-948-564-4	Sequence 4, Appl
39	215	8.0	500	3	US-08-881-784-9	Sequence 9, Appl
40	214.5	8.0	496	3	US-08-881-784-1	Sequence 1, Appl
41	214.5	8.0	496	4	US-09-292-768-2	Sequence 2, Appl
42	214.5	8.0	496	4	US-09-292-768-66	Sequence 66, Appl
43	214	8.0	490	1	US-08-201-118-13	Sequence 13, Appl
44	214	8.0	490	2	US-08-238-821B-13	Sequence 13, Appl
45	214	8.0	490	5	PCT-US95-05744-13	Sequence 13, Appl

## ALIGNMENTS

RESULT 1  
US-08-622-166A-2  
Sequence 2, Application US/08622166A  
Patent No. 5952545  
GENERAL INFORMATION:  
APPLICANT: KONCZ, CSABA  
APPLICANT: MATHUR, JAIDEEP  
APPLICANT: SZEKERES, MIKLOS  
APPLICANT: ALTMANN, THOMAS  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING CYTOCHROME  
TITLE OF INVENTION: P450-TYPE PROTEINS INVOLVED IN THE BRASSINOSTEROID SYNTHESIS  
NUMBER OF SEQUENCES: 4  
CORRESPONDING ADDRESSES:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP  
STREET: P.O. Box 747  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/622,166A  
FILING DATE: 27-MAR-1996  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: SVENSSON, LEONARD R.  
REGISTRATION NUMBER: 30,330  
REFERENCE/DOCKET NUMBER: 0147-0153P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX: 248345  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 472 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-622-166A-2

Query Match 36.5%; Score 978.5; DB 2; Length 472;  
Best Local Similarity 41.9%; Pred. No. 2.1e-89;  
Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;  
QY 12 LLLPLSLLLPLLLKRRNRKTRFNPCKSGMPFLGRTIGYLRKPYTATTLDDFMQKH 71

```

Db 7 LLLLSIAAGFL--LLRRTYRRMGLPPGSLGLPIGFTFOLIGAYKTENPEPFIDER 63
QY 72 VSKYKTYRSLTGEPTIVSADAGLNRFILONEGRLEFECSPKISIGILGKMSLVLYGD 131
Db 64 VARGSVFMTLHFGEPITFSADPTNRFVLONEGKLFECSPASICMLKHSLLKGS 123
QY 132 MHRMRSISLNFSLHARLRTLLKDVRRHFLFVLDNQONSIFSAODEAKKFTFNLMKH 191
Db 124 LHKRHSLLTMSFANSSTIKDHLMDIDLVRFNLDSSRYLL--MEAKKITFELTVKQ 181
QY 192 IMSMDPEEETEOUKKEYVTFMKGVASAPLPGTAVHKAQSATILKFERMEERKL 251
Db 182 LMSFDPG-EMSESLRKEYLVIEGFSLPLPFTSTYRKAIQAR-----RKVAEALT 232
QY 252 DIKEODEEBEVKTEDEAEMSKSDHVRKORTDDLLCWLKHSNLSTEOILDLISLFA 311
Db 233 VVWVKRREEEGAE-----RKMDLAALLAADGFSDEEIVDFVALLVA 278
QY 312 GHETSVAIALAIFLQACPKAVELEREHLEIARAKKEGSELMNDYKKMDFTCQVI 371
Db 279 GYETSTIMTLAVKFLFETPLAQLKEHEHKTAKMSD--SYLEMSDYKSMFTCCV 336
QY 372 NETLRGNVVFELHRAKLDVRYKGYDIPSGKVLPIYSAVHLNDSRYDOPNLEPNRMQ 431
Db 337 NETLRVANIIGVFRAMTDEIKGYKIPKGMKFSSFRVHLDPNHFKDARTENPMRWQ 396
QY 432 QONNGASSSGSFSTYGN--YMPFGGPRLCAGSELAKLEMAVFIHLVLKFMWELAD 490
Db 397 -----SNSVYTPSNVFTPEGGPRLCPCGYELARVALSVFLHRLVTFGSWPAEQ 446
QY 491 DQPAFPFVDFPNGLPIRVS 511
Db 447 DKLVFFPTTRQKRYPIFYKR 467

```

```

RESULT 2
US-08-622-166A-4
; Sequence 4, Application US/08622166A
; Patent No. 5952545
; GENERAL INFORMATION:
; APPLICANT: KONCZ, CSABA
; APPLICANT: MATHUR, JAIDEER
; APPLICANT: SZEKERES, MIKLOS
; APPLICANT: ALTMANN, THOMAS
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING CYTOCHROME
; TITLE OF INVENTION: P450-TYPE PROTEINS INVOLVED IN THE BRASSINOSTEROID SYNTHESIS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/622,166A
; FILING DATE: 27-MAR-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: SVENSSON, LEONARD R.
; REGISTRATION NUMBER: 30,330
; REFERENCE/DOCKET NUMBER: 0147-0153P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 4:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 472 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-622-166A-4

```

```

Query Match 36.5%; Score 978.5; DB 2; Length 472;
Best Local Similarity 41.9%; Pred. No.2,1e-89;
Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

```

```

QY 12 LLLPSLLSLFLILKRRNRKTRFNLPPKSGMPLEGTIGYLYKPYATTAITGLDFMQOH 71
Db 7 LLLLSIAAGFL--LLRRTYRRMGLPPGSLGLPIGFTFOLIGAYKTENPEPFIDER 63
QY 72 VSKYKTYRSLTGEPTIVSADAGLNRFILONEGRLEFECSPKISIGILGKMSLVLYGD 131
Db 64 VARGSVFMTLHFGEPITFSADPTNRFVLONEGKLFECSPASICMLKHSLLKGS 123
QY 132 MHRMRSISLNFSLHARLRTLLKDVRRHFLFVLDNQONSIFSAODEAKKFTFNLMKH 191
Db 124 LHKRHSLLTMSFANSSTIKDHLMDIDLVRFNLDSSRYLL--MEAKKITFELTVKQ 181
QY 192 IMSMDPEEETEOUKKEYVTFMKGVASAPLPGTAVHKAQSATILKFERMEERKL 251
Db 182 LMSFDPG-EMSESLRKEYLVIEGFSLPLPFTSTYRKAIQAR-----RKVAEALT 232
QY 252 DIKEODEEBEVKTEDEAEMSKSDHVRKORTDDLLCWLKHSNLSTEOILDLISLFA 311
Db 233 VVWVKRREEEGAE-----RKMDLAALLAADGFSDEEIVDFVALLVA 278
QY 312 GHETSVAIALAIFLQACPKAVELEREHLEIARAKKEGSELMNDYKKMDFTCQVI 371
Db 279 GYETSTIMTLAVKFLFETPLAQLKEHEHKTAKMSD--SYLEMSDYKSMFTCCV 336
QY 372 NETLRGNVVFELHRAKLDVRYKGYDIPSGKVLPIYSAVHLNDSRYDOPNLEPNRMQ 431
Db 337 NETLRVANIIGVFRAMTDEIKGYKIPKGMKFSSFRVHLDPNHFKDARTENPMRWQ 396
QY 432 QONNGASSSGSFSTYGN--YMPFGGPRLCAGSELAKLEMAVFIHLVLKFMWELAD 490
Db 397 -----SNSVYTPSNVFTPEGGPRLCPCGYELARVALSVFLHRLVTFGSWPAEQ 446
QY 491 DQPAFPFVDFPNGLPIRVS 511
Db 447 DKLVFFPTTRQKRYPIFYKR 467

```

```

RESULT 3
US-08-724-466B-2
; Sequence 2, Application US/08724466B
; Patent No. 6063606
; GENERAL INFORMATION:
; APPLICANT: Beckovitch, P. Martin, White, Jay A.,
; APPLICANT: Beckovitch, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; ZIP: M5L 1A9
; COUNTRY: Canada
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
; COMPUTER: COMPAQ, IBM PC compatible
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,466B
; FILING DATE: October 1, 1996
; PRIOR APPLICATION DATA:

```

```

: APPLICATION NUMBER: 08/667,546
: FILING DATE: June 21, 1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Hunt, John C.
: REGISTRATION NUMBER: 36,424
: REFERENCE/DOCKET NUMBER: 50767/00004
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 863-4344
: TELEFAX: (416) 863-2653
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 492 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-724-466B-2

```

```

Query Match      18.8%; Score 503; DB 3; Length 492;
Best Local Similarity 27.6%; Pred. No. 8,4e-42;
Matches 144; Conservative 95; Mismatches 216; Indels 66; Gaps 14;

```

```

OY 7 HTLLPLLLPSLSLLFL-----ILKRRNRKTRFNLPPKSGMPFGTIGYLPK 58
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4 YTLAVTFELCTIVLLEFLAIAVKLMLMIRVDNCRSPPLPGTMGLPIGTGLTI-- 61
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 59 YTATLIDFMQOHNSKYKIRSNLFGPTIVSADAGLNRFILONEGRLEFCSTYRSIG 118
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 62 ---LQRRKFLMKRQKGCYIKTHLFGNPYRVGADNVRKILGHEHLVSQWPAVSRT 118
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 119 ILGKWSMLVLVGDHNRD-----MRSISLNLFLSHARLRTILKDYERHTLFVLDNW-QONS 172
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 119 ILGSDTLNSVNGVQHKKKKAIMAFSDALEH-----YIPVQGEKSAIQEMLQKDS 172
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 173 IFSADDEKKTFFNLMAKHNSMDPGEET--EOLKKEVTYFMKGVSAPLNPGTAH 230
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 173 CVLVYPEKKKLMFRIAMRILLEGPEQIKTDEOLVEAFEEIMIKLFLPIDVPSGL 232
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 231 ALQSRATILKFERKMERKLDIKEEDQEEEVKTEDEAEKSKSHVAKQRTDDLLGW 290
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 233 GLNRK---NTHSKIEENIRKKITQDDNENEQYKDALQLL---INRSRSD----- 279
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 291 LKHSNLSLEQILDILSLFAGHETSSVAIALAIFFLQACPKAVEELREHELEIRAKE 350
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 280 ----PFSIQAKKEATELLFGGHETATASLVFGLNTEVQKVEB---VQEKVE 331
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 351 LG---ESELNMDYKKMDFTQCYINETLRIGNVVRFLHRAKLDVRYKKGYDIPSGMYL 406
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 332 MGYTPGKGLSMELLDQKTYGCVIKETLRINPVPYGGFRVALKTFEELNGYQIPKGMV 391
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 407 PVISAVHLDNSRYQOPNLFNPMWMOQONNGASSSGSFSFTWGNMYMPGGGPRICASE 466
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 392 YSIDCTHADVAFPNKKEEFQPERF-----SKLEDGSRF--NYIPFGGSGRMCVKE 442
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 467 LAKLEMAVFIHHLVLFKFMWELAEDDQ-----PFAPFVDFP 502
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 443 FAKVLKIFLVEILOHCNWLISNGPPTKGTPTIYVVDNLP 483
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 4
US-08-724-466B-4
: Sequence 4, Application US/08724466B
: Patent No. 6063606
: GENERAL INFORMATION:
: APPLICANT: Petkovich, P. Martin, White, Jay A.,
: APPLICANT: Beckwith, Barbara R., Jones, Glenville
: TITLE OF INVENTION: Retinoid Metabolizing Protein
: NUMBER OF SEQUENCES: 30
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Blake, Cassels & Graydon
: STREET: Box 25, Commerce Court West
: CITY: Toronto
: ZIP: M5L 1A9

```

```

: COUNTRY: Canada
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
: COMPUTER: COMPAQ, IBM PC compatible
: OPERATING SYSTEM: MS-DOS 5.1
: SOFTWARE: WORD PERFECT
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/724,466B
: FILING DATE: October 1, 1996
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: 08/667,546
: FILING DATE: June 21, 1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Hunt, John C.
: REGISTRATION NUMBER: 36,424
: REFERENCE/DOCKET NUMBER: 50767/00004
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 863-4344
: TELEFAX: (416) 863-2653
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 497 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-724-466B-4

```

```

Query Match      16.4%; Score 439; DB 3; Length 497;
Best Local Similarity 27.0%; Pred. No. 2.2e-35;
Matches 142; Conservative 85; Mismatches 226; Indels 72; Gaps 15;

```

```

OY 10 LPULLPLSL-----LSLLFLILK-----RRNRKTRFNLPPKSGMPFGTIGYLPK 57
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 LPALLASLCTFVPLPLLEFLAIAIKLMDLVCSGRDSCALPLPPTMGEPFPGTGL---- 58
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 58 PYTATLIDFMQOHNSKYKIRSNLFGPTIVSADAGLNRFILONEGRLEFCSTYRSIG 117
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 59 -QVLRKRFLOMKRRKRGYIKTHLFGNPYRVGADNVRKILGDDRLVSVMHPASVR 117
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 118 GLGKWSMLVLVGDHNRDMSISLNLFLSHARL---TILKDYERHTLFVLDWQO--NS 172
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 118 TLISGSCISNLHSDSHKQKVIYMAFSDALECYVPVTEEVGSS---LEQWLSGGR 173
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 173 IFSADDEKKTFFNLMAKHNSMDP-----GEEFTEOLKKEVYTMKGVSAPLNPGTA 227
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 174 GLVYPEKKRILMFRIAMRILLEGCEPOLAGDQDSE--QOLVEAFEEIMTRNLFSLPIDVPSG 232
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 228 YKALQSRATILKFERKMERKLDIKEED-----QEEEVKTEDEAEKSKSDHYRKORT 282
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 223 LYRGAKARLILHARLEONIRAKTIGLRASEAGCKDALQILIEHSW-----RGERL 285
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 283 DDDLGLWVILKHSNLSLEQILDILSLFAGHETSSVAIALAIFFLQACPKAVEELREHE 342
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 286 DMQ-----ALKQS--STE-----LFGGHETASATSLTYGLVPHVLOKVBRELK 331
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 343 ELARAKKEELGESELNMDYKKMDFTQCYINETLRIGNVVRFLHRAKLDVRYKKGYDIPSG 402
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 332 SKGLLCKSNODKLMELILQKTYGCVIKETLRINPVPYGGFRVALKTFEELNGYQIPK 391
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 403 WKVLPVISAVHLDNSRYQOPNLFNPMWMOQONNGASSSGSFSFTWGNMYMPGGGPRIC 462
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 392 MWVITSIDCTHDAEFTNKKEEFNDRSAPHPEDASRFS-----FTFPGGGLNSC 442
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 463 AGSELAKLEMAVFIHHLVLFKFMWELAE-----DDQPFAPFVDFP 502
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 443 VGEKFAKILKIFLVEILARHCDWQOLLNGPPTKGTPTIYVVDNLP 487
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 5
US-08-991-677-4
: Sequence 4, Application US/08991677A
: Patent No. 6252135

```

```

: GENERAL INFORMATION:
: APPLICANT: Chiang, Vincent L
: APPLICANT: Carraway, Daniel T
: APPLICANT: Smeltzer, Richard H
: TITLE OF INVENTION: Production of Syngyn1 lignin in Gymnosperms
: FILE REFERENCE: 50617
: CURRENT APPLICATION NUMBER: US/08/991.677A
: CURRENT FILING DATE: 1997-12-16
: EARLIER APPLICATION NUMBER: US 60/033.381
: EARLIER FILING DATE: 1996-12-16
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: Patentia Ver. 2.0
: SEQ ID NO 4
: LENGTH: 511
: TYPE: PRT
: ORGANISM: Liquidambar styraciflua
: US-08-991-677-4

```

```

Query Match      10.8%; Score 289; DB 4; Length 511;
Best Local Similarity 22.0%; Pred. No. 2.3e-20;
Matches 121; Conservative 95; Mismatches 196; Indels 138; Gaps 21;

```

```

QY 6 HHTLLPL-LLPLSLLLFLILKRRNRKTRNLPPGSGMPFLGETIGYLKPYTATTL 64
D6 6 HEALQPLPMLFETIPLLLGLVSR--LRQLPYPPGKGLPIGNML--MMDLTHRGL 62
QY 65 GDFMOHVKYKGIYRSNLFGEPTIVSADAGLNREIIONERLFE-----CSYPR 114
D6 63 AKLAKQ---YGLPLHLMKGLHMAVSTPMARQVLOVDNIFSNRPATTAISLYLYDR 118
QY 115 SIGGILGKWSMLVVGDMHRMRSIS--LNFLSHARLRTILKLDVERHTLFVLDMSQMSI 173
D6 119 AD-----MAFAHYGPPWRQMRKLCVWKLFSSRR-----AESME----- 151
QY 174 FSAODE-----AKFTFNLMKHMMSMDPG-----EETE 203
D6 152 -SVDEVDASAVVNASNIGSTVIGELVFLTKNITVRAFGTISHEDODEFVAILOEFS 210
QY 204 QL-----KKEVTEPMKV---VSAPLNLPGTAYHKALOSRAITILFERKKEERKLDI 255
D6 211 QLFQAFNADITPMKVNPGOSINRLN-----KARALDGFIDKIIDH---IQK 257
QY 256 EDOEEVEKTE-----DEAEMSKSDHVRKQRTDDLLGWLKHSNSTEQIIDLI 305
D6 258 GSKNSEEVDTMVDLLAFYGEAKVSESD-----LQNSIKLTNDNIKAI 303
QY 306 LSLFAGHETSSVAIALAIFLQCPKAVELEBREHLEIARAKKELGSELSNMDYKMD 365
D6 304 MDVNFGEETVVASAIEAMLMKSPEDLKVKQDELAVVGLDRVEER-----DFEKL 358
QY 366 FTQCVINETLGNVVRFLHRKALKDVRKGYDIPSGWKVLPVISAHLDSRYDOPNLF 425
D6 359 YLKCVLEVLRLHPPIPLLHETDAEAVGYIIPAKSRVINACAGRODNSADPDTF 418
QY 426 NPMWMOQONNGASSGSGSEFTWGN--YMPFGGPRLCAGSELAKEMAVFIHHLVLF 483
D6 419 RPSPEFLK-----DGVPDFK--GNPFEPFIFGSGRSCPGMGLVALAETVAHLHCF 469
QY 484 NWELEADDP 493
D6 470 TWELPDGMRP 479

```

```

RESULT 6
US-08-948-564-6
: Sequence 6, Application US/08948564
: Patent No. 6121512
: GENERAL INFORMATION:
: APPLICANT: Siminszky, Balazs
: APPLICANT: Dewey, Ralph E.
: APPLICANT: Cordin, Frederick T.
: TITLE OF INVENTION: No. 6121512el Cytochrome P-450 Constructs and

```

```

: TITLE OF INVENTION: Methods of Producing Herbicide-Resistant Transgenic Plants
: NUMBER OF SEQUENCES: 23
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Virginia C. Bennett
: STREET: PO Box 37428
: CITY: Raleigh
: STATE: No. 6121512th Carolina
: COUNTRY: USA
: ZIP: 27627
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentia Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/948.564
: FILING DATE:
: CLASSIFICATION: 800
: ATTORNEY/AGENT INFORMATION:
: NAME: Bennett, Virginia C.
: REGISTRATION NUMBER: 37,092
: REFERENCE/DOCKET NUMBER: 5051-409
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-854-1401
: TELEFAX: 919-854-1401
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 513 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-948-564-6

```

```

Query Match      10.48%; Score 279.5; DB 3; Length 513;
Best Local Similarity 22.9%; Pred. No. 2.1e-19;
Matches 118; Conservative 82; Mismatches 216; Indels 99; Gaps 17;

```

```

QY 15 LPSLLSLFLILKRRNRKTRNLPPGSGMPFLGETIGYL---KPYTATLQDFMOQH 71
D6 15 LAFITSGILF---FLKQSKSKKRFNLPPOPPMPYVGNLFQVARGSKPFF-----EYVDV 67
QY 72 VSKYKGIYRSNLFGEPTIVSADAGLNREIIONERLFECSYPR----- 114
D6 68 RLRKGSJFTLMKGRITMILIDAKLVHEAMIQGATATRPENPTRIIFSENKFTVNA 127
QY 115 SIGGILGKWSMLVVGDMHRMRSISLNFLSHARLRTILKLDVERHTLFV-----DSW 168
D6 128 TYGPV---WKSJL-----RRNWQNMLSSTRKE--FRSVNDNMADKLINRLKDEAE 173
QY 169 QONSFSAODEAKKFTFNLMKHMMSMDPGEEETEOLKKEVTFPMKGVSA----- 219
D6 174 KNGVWVWKLARAVCIILVACFGLMEDEYVERIDQ---YKSVLITLDRIDYL 229
QY 220 PLNLP--GTAYHKALOSRAITILFERKMEERKLDIKEEDOEVEVKTEDEAEMSKSDH 277
D6 230 PLSPEFSQKQKALEVREQVEFLVPIEQRRRAIQNG-----SDHT 273
QY 278 RKQRTDDL-----GWLKHSNSTEQIIDLISLFFAGHETSSVAIALAIFLQACP 331
D6 274 AYTFSYLDTLFELKVEG---KKSAPSDAELVLSLSEFLNGSTDTTAVAVENGIAQLANP 330
QY 332 KAVELEBREHLEIARAKKELGESELSNMDYKMDFTQCVINETLRLGNVRFLLHRKALKD 391
D6 331 NVQTKLYEE-----IKRYVEKKYDEKDYKMYLIANVVELLRKRPPIHVLHAYTE 384
QY 392 -VRYKGYDIPSGWKVLPVISAHLDSRYDOPNLFNPMWMOQONNGASSGSGSEFTWGN 450
D6 385 PTLGQYDIPIDANVVEYTPAIADDPKMWLNPEKEFDEPRFISGGEADITG-----VTGV 439
QY 451 NYMPFGGPRLCAGSELAKEMAVFIHHLVLFKNW 485
D6 440 KMPFVGVRICPGIAMATYVHIHLMAMRMOVEFEM 474

```



```

RESULT 7
US-09-091-432-2
: Sequence 2, Application US/09091432
: Patent No. 5981837
: GENERAL INFORMATION:
: APPLICANT: Chapple, Clint
: TITLE OF INVENTION: A Method For Regulation Of Plant Lignin Composition
: FILE REFERENCE: 7024-325
: CURRENT APPLICATION NUMBER: US/09/091,432
: EARLIER FILING DATE: 1998-06-18
: EARLIER APPLICATION NUMBER: PCT/US96/20094
: EARLIER FILING DATE: 1996-12-19
: EARLIER APPLICATION NUMBER: US 60/009,119
: EARLIER FILING DATE: 1995-12-22
: EARLIER APPLICATION NUMBER: US 60/013,388
: EARLIER FILING DATE: 1996-03-14
: NUMBER OF SEQ ID NOS: 3
: SOFTWARE: Microsoft Word 2.0C
: SEQ ID NO 2
: LENGTH: 520
: TYPE: PRT
: ORGANISM: Artificial
: FEATURE:
: OTHER INFORMATION: Sequence is deduced from DNA sequence of SEQ ID NO:1
: Patent No. 5981837
: US-09-091-432-2

```

```

Query Match          9.8%; Score 264; DB 2; Length 520;
Best Local Similarity 20.3%; Pred. No. 7.6e-18;
Matches 109; Conservative 103; Mismatches 208; Indels 118; Gaps 18;

```

```

: 11 PULLPSLLSLFLILKRRNRKTRFNLPGKSGMFLGERTGYLKPYATATLGDPMQ 70
: 15 PTLIVIVVSLFISFTRRRRP---YPCGRGMPIIGNL-MMDQTHRGANL--- 67
: 71 HVSKGKTYRSLFEGEPTIVSADAGLNRFILONEGRLE-----CSTPRSIGIL 120
: 68 -ARKYGGCHLMGFLAHYAVSSPEVAROVLOVDSPSNPATIAISLYTRAD--- 122
: 121 GKMSLVLVGDMHRMRSIS-LNFLSHARLRTILKDVVERHTLFVLDSMQNSISADE 179
: 123 ---MAFAHYGPRWKRKVCVKVSRKR-----AESVA---SVRDE 158
: 180 AKKFTFNL-----MAKHISM-----DPGEETEOLKKEVTFMKGV 216
: 159 VDKWVRSVSCVNGKPIVNGEOLFALTRNITYRAFGSACENKODEFIRLOEF-SKLFCA 217
: 217 VSAPLNDGTAY-----HKAOSRAITLFIKRMKEERKLDIKEEDOEVEEVKTE-DE 268
: 218 FNVADFIFYFGVIDPQINKRLVKARNLDGFIID-----DIIDEHMKKENONAVDD 269
: 269 AEMSSSDVHRKQRTDDLLGLWVLKHSNSTE-----QIDLLISLLRAGHET 315
: 270 GGVVDTDWV-----DLLAEYSEPAKLVSTADLONSIKILTRDNKAIIMVMGGET 323
: 316 SSVAIALAIFFOACPKAVVEELREELHETARAKELGESELMWDYKKMDFQCYINETL 375
: 324 VASALEMALTLLRSPEDLKRQOGLAEVVGIDRVEES-----DIEKITYLKJLKETL 378
: 376 RLGNVVRFLHRRKALDKVRYKGYDIPSGMKVLFVISAVHLDNSRYDOPNLFNWMQOONN 435
: 379 RMHNPRIPLLHETADTIDGFFIFPKSRVMINAFIGRDSWTDPTDFRSPRELE--- 435
: 436 GASSSGSSSEFTWGNWVMPFGGGRPLCAGSELAKEAMAVFIHLVLKFNWELAEEDOP 493
: 436 ---PGVDFEFGSNFEFTIPFGSGRSCPMOGLYALDLAAVAHILHCFWKLDPGMR 489

```

RESULT 8  
US-08-457-274A-25

```

: Sequence 25, Application US/08457274A
: Patent No. 5734086
: GENERAL INFORMATION:
: APPLICANT: Scott, Jeffrey G.
: APPLICANT: Tomita, Takashi
: TITLE OF INVENTION: Cytochrome P450lpr Gene and Its Uses
: NUMBER OF SEQUENCES: 29
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Nixon, Hargrave, Devans & Doyle
: STREET: P.O. Box 1051, Clinton Square
: CITY: Rochester
: STATE: New York
: COUNTRY: USA
: ZIP: 14603
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/457,274A
: FILING DATE:
: CLASSIFICATION: 800
: ATTORNEY/AGENT INFORMATION:
: NAME: Goldman, Michael L.
: REGISTRATION NUMBER: 30,727
: REFERENCE/DOCKET NUMBER: 19603/240 (D-1519)
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 716-263-1600
: TELEFAX: 716-263-1304
: INFORMATION FOR SEQ ID NO: 25:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 504 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Rat
: STRAIN:
: DEVELOPMENTAL STAGE: Adult
: POSITION IN GENOME:
: CHROMOSOME/SEGMENT:
: US-08-457-274A-25

```

```

Query Match          9.8%; Score 263.5; DB 1; Length 504;
Best Local Similarity 24.1%; Pred. No. 8.1e-18;
Matches 132; Conservative 89; Mismatches 213; Indels 113; Gaps 25;

```

```

: 9 LPLLLPSLLSLFLILKRRNRKTR--FNLPGKSGMFLGERTGYLKPYATATL 64
: 3 LLSALTLEWVLLAVLVLLYFGGTRTHGLFKKQIGPKRPLRFGTYLVNYT-----MGL 57
: 65 GDFMOOHVSKYKTYRSLNF-GE-PTIVSADAGLNRFILONEGRLEEC-----SYPRSIG- 117
: 58 WFEVDECHKKYKIV--GLFDGOMPLFALTIDTEMKIVNL-----YECPSVFTNRDFFP 110
: 118 -GILKMSLVLVGDMHRMRSISLNFLSHARLRTILKDVVERHTLFVLDSMQNS---- 172
: 111 VQIMKRAVAVANDEWKRKRALLSPFTFS-GRLEK-MEPIIIOYGDILVKYLKQEAETGK 168
: 173 -----IFGA--QDBAKKFTFNLMAKHISM-MDPGEETEOLK--EYTFMKGVVASPL 221
: 169 PYTMKKVFCAYMDVITSTSGVNVNDSLNNPKDPFEKTKKLLRDFPRLPLSVLFPF 228
: 222 NLPGTAYHKAAL-----QSRATILFIKRMKEERKLDIKEEDOEVEEVKTEDEAEMSKS 274
: 229 LTP---IYEMLNICMPKDSIEFFKFFYR--MKETRL----- 261
: 275 DNVHRKQRTDDDLLGLWV-----KHSNSTEQIIDLLISLFLRAGHETSSVAIALAI 324

```

```

Db 262 DSVOKHRV--DFLOLUMNANHDSKDKESHITALSMEITAQSIIFIFAGYEPTSTLSFVL 319
QY 325 FFLQACRAVELEEBHELEIARAKKEGESELNDYKKMFTOCVINETRLGNVAF 384
Db 320 HSLATHPTOKKLOE--EIDRALP--NKAPPTDYVMEYELMVLNETRLVPIGNRL 374
QY 385 HRKALKDVRKYGYDIPSGWKVLPVISAHLNDSRYDQPNLFNPMRMOQONNGASSGSGS 444
Db 375 ERVCKKDVINGVMPKSGVYMIPSYALHRDPQHMPPEEERPERFSKENGSDIP---- 430
QY 445 FSTGNNYMPFGGPRCAGSELAKLEMAVFIHHLVLFKNELAEDDQPFAPFVDPENG 504
Db 431 -----YVLPFGNGPRNCIGRFAIMNMKALTKVLONFSPCKETO----- 473
QY 505 LPIRVSR 511
Db 474 IPLKLSR 480

```

## RESULT 9

PCT-US95-05758-25

```

Sequence 25, Application PC/TUS9505758
GENERAL INFORMATION:
APPLICANT: Cornell Research Foundation, Inc.
TITLE OF INVENTION: Cytochrome P4501pr Gene and Its
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle
STREET: P.O. Box 1051, Clinton Square
CITY: Rochester
STATE: New York
COUNTRY: USA
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/05758
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
TELECOMMUNICATION INFORMATION:
REFERENCE/DOCKET NUMBER: 19603/241 (D-1519)
TELEPHONE: 716-263-1304
TELEFAX: 716-263-1600
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 504 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Rat
STRAIN: Unknown
DEVELOPMENTAL STAGE: Adult
POSITION IN GENOME:
CHROMOSOME/SEGMENT: Unknown
PCT-US95-05758-25

```

Query Match 9 88; Score 263.5; DB 5; Length 504;  
 Best Local Similarity 24.18; Pred. No. 8.1e-18;  
 Matches 132; Conservative 89; Mismatches 213; Indels 113; Gaps 25;

```

QY 9 LPLLLPSSLTLLFLILKRNRKTR--FNLP--DGKSGMPELGETIGYLRKYATATL 64
Db 3 LLSALTFETWLLAVLVLLFGFGRTHGLFKKGIPOKPLPFPGYLVNY-----MGL 57
QY 65 GDMQOHVSKTKGKIYRSNLF-GE-PTIVSADAGLNRLTONEGRLEFC---STPRSTG- 117
Db 58 WKFDVECHKKKYKTIW--GLFDGQMPFLAITDTEMIKNVL-----VKECFSYFTNRROFGP 110
QY 118 -GIIGKMSMLVLVDGMHDMNISLNFSLHARLRLTLKDVERTHLPLVDSMOONS----- 172
Db 111 VLMGKAVSAKDEWKRYRALSPFTTS-GRLE-MPPIEQYGDILLVTKLQDAETGK 168
QY 173 -----IFSA--ODEAKKFTPLMAKHIMS-MDGEETEOLEK--EYVTPMKGVSAPL 221
Db 169 PVTMKVGEAVSMQVITSTSGVAVDLSNMPKDPFEVETKLLRFPDFDLPLSVLFPF 228
QY 222 NLPGTAHKKAL-----QSRATILKFTERMEERKLDIKEEDQEBEERVTEDAEENKS 274
Db 229 LTP---IYEMLNICMPKDSIEFFKKFYVR-MKETRL----- 261
QY 275 DHVRKORTDDDLGWL-----KHSNLSSTEQILDLTLLSPAGHETSVAIALAI 324
Db 262 DSVOKHRV--DFLOLUMNANHDSKDKESHITALSMEITAQSIIFIFAGYEPTSTLSFVL 319
QY 325 FFLQACRAVELEEBHELEIARAKKEGESELNDYKKMFTOCVINETRLGNVAF 384
Db 320 HSLATHPTOKKLOE--EIDRALP--NKAPPTDYVMEYELMVLNETRLVPIGNRL 374
QY 385 HRKALKDVRKYGYDIPSGWKVLPVISAHLNDSRYDQPNLFNPMRMOQONNGASSGSGS 444
Db 375 ERVCKKDVINGVMPKSGVYMIPSYALHRDPQHMPPEEERPERFSKENGSDIP---- 430
QY 445 FSTGNNYMPFGGPRCAGSELAKLEMAVFIHHLVLFKNELAEDDQPFAPFVDPENG 504
Db 431 -----YVLPFGNGPRNCIGRFAIMNMKALTKVLONFSPCKETO----- 473
QY 505 LPIRVSR 511
Db 474 IPLKLSR 480

```

## RESULT 10

US-08-313-075A-30

```

Sequence 30, Application US/08313075A
Patent No. 5639870
GENERAL INFORMATION:
APPLICANT: Holton, Timothy A.
APPLICANT: Cornish, Edwin C.
APPLICANT: Tanaka, Yoshikazu
TITLE OF INVENTION: GENETIC SEQUENCES ENCODING FLAVONOID
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: U.S.A.
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/313,075A
FILING DATE: 30-NOV-1994
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PL 1538/92
FILING DATE: 27-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PL 6698/93

```

FILING DATE: 07-JAN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: AU PCT/AU93/00127  
FILING DATE: 25-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Digilio, Frank S.  
REGISTRATION NUMBER: 31,346  
REFERENCE/DOCKET NUMBER: 9433  
TELEPHONE: (516) 742-4343  
TELEFAX: (516) 742-4366  
TELEX: 230 901 SANS UR  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 476 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-313-075A-30

Query Match 9.6%; Score 257; DB 1; Length 476;  
Best Local Similarity 24.0%; Pred. No. 3.3e-17;  
Matches 123; Conservative 87; Mismatches 227; Indels 76; Gaps 21;

QY 12 LLLPSSLILLLFLILKRRNRKTRFNLPPKSGMPFLGRTIGYIKPTATTLGDMQOH 71  
DB 9 LLLPSSLILLLFLILKRRNRKTRFNLPPKSGMPFLGRTIGYIKPTATTLGDMQOH 59  
QY 72 VSKYGIYRSNLFGEPTIVSADAGLNRFILONEGRLEFCSY-PRSIGILGKMSLVLV- 129  
DB 60 AKKGPIHNDLGKSAVYVSPEAKRVLTHTDLAF--AKRPKLGLIYICYNSSDLAF 117  
QY 130 ---GDMHDMRSIS-LNFLSHARLTILKDYERHTFLVDSMOONS-----IFS 175  
DB 118 SPYGDYWMQMRKICVLEYSKAVNRS--FNSIRDEILIMIDFLSSSGKPVNITERIFS 175  
QY 176 AQDEAKKRTFLMKNHMSMDPGE-EETDQKKEVTFMKGVNSAPLPLPGAV-HKMLQ 233  
DB 176 -----FTSMICKSVGKRIKEDDCIRHVKKWTGLIDGDVADI-FPSIRFLHVLIG 227  
QY 234 SRATLKEFERKMEERKLDIKEEDQEEVEKTEDEAKSKSDHVRKQRTDDLLGWLKH 293  
DB 228 MKGRIMD-VHRRV-----DAIVEVNNHK-ETLRGTGTGEVGEEDLIVLRL 275  
QY 294 SN-----LSTEDLDLILSLPAGHETSSVALALAIFFLOACPKAVEELREHLEIAR 346  
DB 276 KEEGDLPRTNDNIKAIFNDMFAAGTETSTTIMWAVEIMKNSVFAKQAQAEVREYFK 335  
QY 347 AKKELGESELMWMDYKKMDPQOCVINTLRL-GNVYRFLHRLKALKDYKKGIDIPSGKV 405  
DB 336 GKEFFDE---DDEELNLYKLVIRETLRLPLPLLPRECRRETINGTTPPLNTRY 390  
QY 406 LPVISAHLNDSRYDOPMLFNPWRMOQONNGASSSGSFSFTGNN--YMPFGGPRICA 463  
DB 391 IVNWVAIGRODKYMDADASFPEREH-----NSLMPAGNFFYLPRGSRRLCP 440  
QY 464 GSELAKLEMAVFIHHLVLKFWELAEEDQPAF 496  
DB 441 GISFGIANVYHPLAQLLHYFDMRLPTGVDPRDF 473

RESULT 11  
US-08-201-118-7  
Sequence 7, Application US/08201118  
Patent No. 5786191  
GENERAL INFORMATION:  
APPLICANT: GOLDSTEIN, Joyce A.  
APPLICANT: ROMKES-SPARKS, Marjorie  
TITLE OF INVENTION: CLONING AND EXPRESSION OF COMPLEMENTARY  
TITLE OF INVENTION: DNAS FOR MULTIPLE MEMBERS OF THE HUMAN CYTOCHROME P450 2C  
SUBFAMILY  
NUMBER OF SEQUENCES: 44

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Kourie and Crew  
STREET: 379 Lytton Avenue  
CITY: Palo Alto  
STATE: California  
COUNTRY: US  
ZIP: 94301  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/201,118  
FILING DATE: 22-FEB-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/864,962  
FILING DATE: 09-APR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Liedeschuetz, Joe  
REGISTRATION NUMBER: 37,505  
REFERENCE/DOCKET NUMBER: 15280-192-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 326-2400  
TELEFAX: (415) 326-2422  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 490 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-201-118-7

Query Match 9.6%; Score 256.5; DB 1; Length 490;  
Best Local Similarity 20.8%; Pred. No. 3.9e-17;  
Matches 110; Conservative 99; Mismatches 202; Indels 119; Gaps 20;

QY 11 LLLPSSLILLLFLILKRRNRKTRFNLPPKSGMPFLGRTIGYIKPTATTLG 65  
DB 3 PNVVLVLTSLFSLFSLNRQSCRK--LPQPTPLPIIGNMLOIDVNDICKSFT----- 55  
QY 66 DPMQVSK-YGIYRSNLFGEPTIVSADAGLNRFILONEGRLEFCSYPRSIGILGK-W 123  
DB 56 -----NFSKVVGPVTVFVGMNPIVFFHGYEAVKDALIDNGEEFSGRGNPSISQRTKGL 110  
QY 124 SMLVVGDMHDMRSISLNFLSHARLTILKDYERHTFLVDSMOONS-----FVL 165  
DB 111 GIISSNGRMRKEIRRFSLNLFNFGMKRSIEDRYOEENHCLVEELKRTKASPCDPTFL 170  
QY 166 DSMQONSIFSADEAKKRTFLMKNHMSMDPGEETEOLKKEVTFMKGVNSAPLPLPGAV-HKMLQ 233  
DB 171 GCAPCNVICSVYQ-KRDPYK-----DQNFLLMKRFENRRLNS 210  
QY 220 PL-----NLPGTAHKAQSRATILKTERKMEERKLDIKEEDQEEB---V 263  
DB 211 PWIQYCNPPRLIDCFPGH-HNKVLKNAVALTRSYIREYVKEHQASLDVNNPRDFMDFCL 269  
QY 264 KTEDEAKSKSDHVRKQRTDDLLGWLKHSNLSSTEDLDLILSLPAGHETSSVALALAI 323  
DB 270 KMEDEKDKOSE-----FNTEMLVGTADLPVAGTETSTTLRYG 309  
QY 324 IFFLOACPKAVEELREHLEIARAKKELGESELMWMDYKKMDPQOCVINTLRLGNNV 381  
DB 310 LLLLKHPVLAQVQELIDNV-IGRHRSPC-----MDRSMRPYTDVAVNHDIQYSDLV 362  
QY 382 -RFLRKALKDYKKGIDIPSGKVLPAVISAHLNDSRYDOPMLFNPWRMOQONNGASS 440  
DB 363 PTGVHAAVTTDKFRNYILIPKGTITMALITSVLHDDKEPNNPNDIPDFHLDKN----- 416  
QY 441 GGSFSFTGNNNMPFGGGRICAGSELAKLEMAVFIHHLVLKFWELAEED 490

Db 417 --GNFKK-SDYFMPFSAGKRICAGEGLARMELFLFTIILQNFNLKSVDD 463

## RESULT 12

US-08-238-821B-7  
Sequence 7, Application US/08238821B  
Patent No. 5912120  
GENERAL INFORMATION:  
APPLICANT: GOLDSTEIN, Joyce A.  
APPLICANT: ROMKES-SPARKS, Marjorie  
APPLICANT: DE MORAIS, Sonia M.F.  
TITLE OF INVENTION: CLONING, EXPRESSION AND DIAGNOSIS OF HUMAN  
TITLE OF INVENTION: CYTOCHROME P450 2C19: THE PRINCIPAL DETERMINANT OF S-  
NUMBER OF SEQUENCES: 61  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: US  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/238,821B  
FILING DATE: 06-MAY-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/201,118  
FILING DATE: 22-FEB-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/864,962  
FILING DATE: 09-APR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Liebeschuetz, Joe  
REGISTRATION NUMBER: 37,505  
REFERENCE/DOCKET NUMBER: 15280-192110US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 326-2400  
TELEFAX: (650) 326-2422  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 490 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-238-821B-7

Query Match 9.6%; Score 256.5; DB 2: Length 490;  
Best Local Similarity 20.8%; Pred. No. 3,9e-17;  
Matches 110; Conservative 99; Mismatches 202; Indels 119; Gaps 20;

QY 11 PLLLLSLSLFLFLILKRRNRKTRFNLPPGKSGWPLGFTI-----GYLKPYATTATLG 65  
DB 3 PFVAVLVLCLSPMLFLSLMRGCRRK--LPPGPTPLPIGMLQIDVKDICKST----- 55  
QY 66 DFMQOHYSK-IGKTYRNSLGEPTIVSADAGLNFILONBSRLFECSYPSISGILGK-W 123  
DB 56 -----NESKYVGPVETYYFGNRPVVFHGYEAVKALIDNGEERSGNGNSPISQITKGL 110  
QY 124 SMLVYGDMDHMDSISLNFSLHARLRTIILKD---VERHTL-----FVL 165  
DB 111 GIISSNGKRKKEIRFSITLNRNGMGKRSIEDRYQDEAHCLVLELKRKTASPCDPPFILL 170  
QY 166 DSMOONSIFSADBAKRTFNLMKAIMSMDPGEEETEOLKKEYVTMK-----GVVSA 219

Db 171 GCAPCNVICSVFO-KREDYK-----DONFLTLMKRENNFRILNS 210  
QY 220 PL-----NLPTAHKALQSRATILKFIERKMEERKIDIKEEDEDEEE-----V 263  
DB 211 FMIQCNPNFLLIDCFPT-HNKVKLVNLTSTRYSIREKVKHQASLDVNNRDRMDCFLI 269  
QY 264 KTEDEAEKSKSDHYRKQRTDDLLGWVLKHSNLSTEOILDLILSLFAGHETSSVAIALA 323  
DB 270 KMEDEKDNQSE-----FNIENLVGVALDLFAGTTTSTTLRYG 309  
QY 324 IFELQACRAVEELREE--HLEIARAKKEGSELNMDYKMKDFQCVINETLRGNV 381  
DB 310 LLLILKHEVYAKVQOEELIDHV-IGRHRSPC-----MODRSHMPTDAVHIEIORSDLV 362  
QY 382 -RFLHRAKLVKRVKGYDIPESGWVLPYISAVHLDNSRYDDPNLFNPRMOOONNGASS 440  
DB 363 PTGVPNAVTTDTKFRNRYLIPGTTIMALLTSLVLDKKEFPNPIFDQHPILDKN----- 416  
QY 441 GSGSFSTWGNVMPFGGPRLCAGSELAKLEMAVEIHHLVLKFWMLAED 490  
DB 417 --GNFKK-SDYFMPFSAGKRICAGEGLARMELFLFTIILQNFNLKSVDD 463

## RESULT 13

PCT-US95-05744-7  
Sequence 7, Application PC/TUS9505744  
GENERAL INFORMATION:  
APPLICANT: GOLDSTEIN, Joyce A.  
APPLICANT: ROMKES-SPARKS, Marjorie  
APPLICANT: DE MORAIS, Sonia M.F.  
TITLE OF INVENTION: CLONING, EXPRESSION AND DIAGNOSIS OF HUMAN  
TITLE OF INVENTION: CYTOCHROME P450 2C19: THE PRINCIPAL DETERMINANT  
NUMBER OF SEQUENCES: 61  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew  
STREET: 379 Lytton Avenue  
CITY: Palo Alto  
STATE: California  
COUNTRY: US  
ZIP: 94301  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/05744  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/238,821  
FILING DATE: 06-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/201,118  
FILING DATE: 22-FEB-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/864,962  
FILING DATE: 09-APR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Dow, Karen B.  
REGISTRATION NUMBER: 29,684  
REFERENCE/DOCKET NUMBER: 15280-192-1-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 326-2400  
TELEFAX: (415) 326-2422  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 490 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear



MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/948,564  
FILING DATE:  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Bennett, Virginia C.  
REGISTRATION NUMBER: 37,092  
REFERENCE/DOCKET NUMBER: 5051-409  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-854-1400  
TELEFAX: 919-854-1401  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 576 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-948-564-16

Query Match 9.1%; Score 244.5; DB 3; Length 576;  
Best Local Similarity 22.7%; Pred. No. 7.9e-16;  
Matches 111; Conservative 91; Mismatches 181; Indels 107; Gaps 22;

QY 63 TLGDMQOHVSKYKITYSNLFGPTIVSADAGLNRFILQ-----NGRLFECSYPRSI 116  
DB 97 SLYDMFLFH---GAVYKLAFCGPKAFVVSPIVAHILRENAFSYDKGLADILEP--- 149  
QY 117 GGIGCK-----WSM--LVLVGDMHDMRSISLNFSLHARLTIL--KDERHRLF 163  
DB 150 --IMCKGLIPADLTWKQRRIYAPAFHNSYLEAMVKIFTTCSERTILKFNKLLGEGCYD 207  
QY 164 VLDWMQONSIFSADDEAKKFTFNLMAKHIMSDPGE--EETDLKKEYVTFMKGVASAPL 221  
DB 208 GPDSIE---LDLEAFSSALDITGLGVFNDFGSVTKESPVIKAVYGTLEAEHRSTF 263  
QY 222 NLPGAVYKALQSRTILKFTERRKKEERKLDK-----EEDQEEBEVKTEDE 268  
DB 264 YIP--YKRIPLARVIYPR--QRKFQD--DLKVINTCGLDLIRNAKESRQETDVE--- 311  
QY 269 AEMSKSDHVRKQRTDDDLGWN--LKHNSLSTEOITLILSLFAGHETSVAIALAIF 326  
DB 312 -KLOQRDYLNLK--DASLLRFLVDMRGADVDDRLDMLMTLNLGHETTAAVLTMAYFL 368  
QY 327 LQACRAVEELREHLEIETARRAKKEIGSELNMDYKKMDFTCVYNETLRLGNVVRFLHR 386  
DB 369 LAQNPSSKKKAQ-----AEVDLVLTGTGRPTFESLKEIYIRLIYVEALRLYPPPLIR 422  
QY 387 KALK-DV-----RYKGYDIPSGMKVLPVISAHLNDSRY--DQNLFPNPRMQQNNNG 436  
DB 423 RSLKSDVLPGGHKGKGDYALPAGTDVF--ISVYNLHRSPLYFMDRPPDDEPERFLVQKN 480  
QY 437 ASSSGSGSFTWGN-----NYMPFGGPRLCAGSELAKLEMAVFIH 477  
DB 481 EEIEG-----WAGLDPRSPRGALYPNNEVISDFALFLPRGGPRKCVGDPQFALMESTVALT 534  
QY 478 HLVLKENMEL 487  
DB 535 MLQNFDEL 544

Search completed: October 17, 2001, 22:07:08  
Job time: 10067 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2001, 19:51:16 ; Search time 41.56 Seconds  
(without alignments)  
940.268 Million cell updates/sec

Title: US-09-502-426-2

Sequence: 1 MFETENHTLLPLLLPLSLLS.....FAFPFVDFPGLPIRYSRL 513

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

PIR\_68: \*  
1: PIR1: \*  
2: PIR2: \*  
3: PIR3: \*  
4: PIR4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2677	99.9	513	2 T46143	steroid 22-alpha-h
2	978.5	36.5	472	1 S55379	cytochrome P450 Cy
3	866	32.3	512	2 H96759	probable steroid 2
4	860	32.1	457	2 D85429	cytochrome P450 1i
5	815	30.4	382	2 T48613	hypothetical prote
6	761.5	28.4	464	2 T07859	cytochrome P450 ho
7	632	23.6	457	2 T04444	cytochrome P450 -
8	611	22.8	482	2 T02739	probable cytochrom
9	581	21.7	487	1 C71417	cytochrome P450 d1
10	579.5	21.6	485	2 A84859	probable cytochrom
11	579.5	21.6	490	2 H86185	hypothetical prote
12	577	21.5	455	2 T48973	cytochrome P450-1i
13	547.5	20.4	489	2 B84733	probable cytochrom
14	497.5	18.6	460	2 D96813	hypothetical prote
15	491.5	18.3	519	1 T02263	cytochrome P450 DW
16	489.5	18.3	255	2 T04602	cytochrome P450 ho
17	479.5	17.9	444	1 S75761	cytochrome P450 -
18	369	13.8	349	2 A86329	hypothetical prote
19	357	13.3	518	2 T20908	hypothetical prote
20	338	12.6	453	2 C83722	cytochrome P450 hy
21	329.5	12.3	517	2 T20907	hypothetical prote
22	324.5	12.1	520	2 T24778	hypothetical prote
23	321.5	12.0	518	2 T24783	hypothetical prote
24	316	11.8	500	2 T04737	cytochrome P450 ho
25	310.5	11.6	520	2 T24777	hypothetical prote
26	310	11.6	500	2 T52175	cytochrome P450 mo
27	306.5	11.4	504	2 A25222	cytochrome P450 3A
28	302.5	11.3	491	2 S31277	cytochrome P450 2B
29	300	11.2	491	2 I84735	testosterone 16a-h

30	299.5	11.2	492	2 S27160	cytochrome P450 2B
31	297	11.1	491	2 A31047	testosterone 16alp
32	297	11.1	494	2 A33293	cytochrome P450 2A
33	294.5	11.0	491	1 O4RBPc	cytochrome P450 2B
34	294	11.0	506	2 D96672	probable Cytochrom
35	294	11.0	516	2 T48140	flavonoid 3',5'-hy
36	292	10.9	491	1 O4RBPc	cytochrome P450 2B
37	291.5	10.9	491	2 S35666	cytochrome P450 2B
38	291.5	10.9	530	1 A27491	lanosterol 14alpha
39	290.5	10.8	502	2 T05246	cytochrome P450, g
40	288.5	10.8	504	2 A29410	cytochrome P450, g
41	287.5	10.7	520	2 T24780	hypothetical prote
42	287	10.7	503	2 JC4702	cytochrome P450 3A
43	284.5	10.6	505	1 S38534	cytochrome P450 76
44	284	10.6	501	2 T04735	cytochrome P450 ho
45	282	10.5	491	2 I49625	testosterone 16a-h

## ALIGNMENTS

RESULT 1  
T46143  
steroid 22-alpha-hydroxylase (DWF4) - Arabidopsis thaliana  
N:Alternate names: protein T3A5.40  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 16-Feb-2001  
R:Biocker, H.; Mewes, H.W.; Mayer, K.F.X.; Lemcke, K.; Schueller, C.; Quetier, F.; S  
submitted to the Protein Sequence Database, December 1999  
C:Accession: T46143  
A:Reference number: Z23024  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-513 <BLO>  
A:Cross-references: EMBL:AL132979  
A:Experimental source: cultivar Columbia; BAC clone T3A5  
C:Genetics:  
A:Map position: 3  
A:Introns: 74/2; 182/3; 233/3; 338/3; 369/3; 396/1; 432/3  
A:Note: T3A5.40  
C:Superfamily: Synecocystis cytochrome P450 slr0574; cytochrome P450 homology  
F:Keywords: heme; iron; metalloprotein  
F:308-484/Domain: cytochrome P450 homology <P45>  
F:462/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match	99.9%	Score 2677	DB 2	Length 513
Best Local Similarity	99.8%	Pred. No. 9.2e-166		
Matches 512	Conservative 1	Mismatches 0	Indels 0	Gaps 0
QY	1	MEFENHTLLPLLLPSLLSLLFLTLKRRNRKTRFNLPPKSGMPFLGFTGYLKPVT	60	
DB	1	MEFENHTLLPLLLPSLLSLLFLTLKRRNRKTRFNLPPKSGMPFLGFTGYLKPVT	60	
QY	61	ATTLDPMQOHTSKGKTYRSLNFGPTIVSADAGLNRIITLNEGRIFECSTPSIGIL	120	
DB	61	ATTLDPMQOHTSKGKTYRSLNFGPTIVSADAGLNRIITLNEGRIFECSTPSIGIL	120	
QY	121	GKSMILVLVGDHMRDRSTSLNFIASHARLTILKDVENHTFLVLDPMQONSIFSADDEA	180	
DB	121	GKSMILVLVGDHMRDRSTSLNFIASHARLTILKDVENHTFLVLDPMQONSIFSADDEA	180	
QY	181	KFTFNLAHKHIMSMDPGEETEOLKKEVVTMGKGVSAPLMPTAYHAKALQSRATILK	240	
DB	181	KFTFNLAHKHIMSMDPGEETEOLKKEVVTMGKGVSAPLMPTAYHAKALQSRATILK	240	
QY	241	FIERMEERKILDKEDDEEVEKTEDEAEKMSQHVRRQRDDDLGVLKHSNLSTEQ	300	
DB	241	FIERMEERKILDKEDDEEVEKTEDEAEKMSQHVRRQRDDDLGVLKHSNLSTEQ	300	
QY	301	ILDILSLFAGHETSVAIALAIFLQACPRAVELREHLEIARAKKEGESELMND	360	

Db 301 ILDLISLFLAGHETSSVAIALAIFLQACPKAVEELREEHLEIARAKKEGESELNMD 360  
 QY 361 YKKMDFTCVINETLRLGNVVFRLHKKALKDVRKYGYDIPSGMKVLPYISAVHLDNSRD 420  
 Db 361 YKKMDFTCVINETLRLGNVVFRLHKKALKDVRKYGYDIPSGMKVLPYISAVHLDNSRD 420  
 QY 421 QPNLFNPMRMOQONNGASSSGSFGSTWGNVMPFGGPRLCAGSELALEMVFIHLV 480  
 Db 421 QPNLFNPMRMOQONNGASSSGSFGSTWGNVMPFGGPRLCAGSELALEMVFIHLV 480  
 QY 481 LKFMNELAEDDOPFAFPVDPNGLPIRVSR 513  
 Db 481 LKFMNELAEDDOPFAFPVDPNGLPIRVSR 513

## RESULT 2

S55379  
 cytochrome P450 CYP90 - Arabidopsis thaliana  
 N:Contains: oxidoreductase (EC 1.-.-.-)  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000  
 C:Accession: S55379  
 R:Sequences: M.: Nemeth, K.; Koncz, Z.; Nagy, F.; Koncz, C.  
 submitted to the EMBL Data Library, May 1995  
 A:Accession: S55379  
 A:Molecule type: mRNA  
 A:Residues: 1-472 <SIZE>  
 A:Cross-references: EMBL:X87367; NID:9853718; PIDN:CAA60793.1; PID:9853719  
 C:Genetics:  
 A:Gene: CYP90  
 C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology  
 C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenase;  
 F:275-440/domain: cytochrome P450 homology <P45>  
 F:418/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 36.5%; Score 978.5; DB 1; Length 472;  
 Best Local Similarity 41.9%; Pred. No. 5.9e-56;  
 Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

QY 12 LLLPSLSLLFLILKRRNRKTRFNLPRPGSGMPLFGETIGYLYKPYATTATLGDPMOCH 71  
 Db 7 LLLSSIAAGFL--LLLRTRRYRMGLRPGSLGLPGLIGFOLIGAYKTENPEPFIDER 63  
 QY 72 VSKYGIYRSNLFGPPTVSADAGLNRFILONEGRLEPCSPRISGILGKMSMVLVIGD 131  
 Db 64 VARGSVYMTLFGEPITFSDAPETNRKRVLDNEGKLEFCSTPASTICNLGKHSILMKGS 123  
 QY 132 MHRMRSISLNLFLSHARLTILKDVERRHTLFVLDMSQONSIFSAODEAKKFTENLMAKH 191  
 Db 124 LHKRHSILTMSFANSSIKDHLMDIDRLVAFNLDSSSVLL--MEKAKITTEFLYKQ 181  
 QY 192 IMSDPRGEELQKKEVYFMKGVASAPLNLPGTAHYKALQSRAATILKTERKMEERKL 251  
 Db 182 LMSDPG--EMSESRKEKELVTEGFFSLPLRFSTYRKAIQAR-----RKVAEALT 232  
 QY 252 DIKEDEDEEVEKTEDEAKMSDHYVRKQRTDDLLGVTLKHSNLSTEQIIDLISLFLA 311  
 Db 233 VVAKKREEEEGAE-----RKMDLALLAADGSDDEIVDFVALLYA 278  
 QY 312 GHETSSVAIALAIFLQACPKAVEELREEHLEIARAKKEGESELNMDYKKMDFTCV 371  
 Db 279 GYETSTIMTLAVFLFETPLALQKKEHEKIRAKMSD--SYSEMSDYKSMFETQCV 336  
 QY 372 NETLRLGNVVFRLHKKALKDVRKYGYDIPSGMKVLPYISAVHLDNSRYDQPNLFNPMRMO 431  
 Db 337 NETLRLGNVVFRLHKKALKDVRKYGYDIPSGMKVLPYISAVHLDNSRYDQPNLFNPMRMO 431  
 QY 432 QONNGASSSGSFGSTWGNVMPFGGPRLCAGSELALEMVFIHLVLFKFMNELAED 490  
 Db 397 -----SNSVTTGSPNVETFPFGGPRLCAGSELALEMVFIHLVLFKFMNELAED 446

QY 491 DOPFAFPVDPNGLPIRVSR 511  
 Db 447 DKLVEPPTTRQKRYPIFKR 467

## RESULT 3

H96759  
 Probable steroid 22-alpha-hydroxylase T9L24.44 [Imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: H96759  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,  
 ansen, N.F.; Hughes, B.; Huizart, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzla  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719  
 A:Accession: H96759  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-512 <STO>  
 A:Cross-references: GB:AE05173; NID:911120803; PIDN:AAG30983.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: T9L24.44  
 A:Map position: 1

Query Match 33.3%; Score 866; DB 2; Length 512;  
 Best Local Similarity 33.8%; Pred. No. 1.2e-48;  
 Matches 181; Conservative 103; Mismatches 173; Indels 76; Gaps 7;

QY 12 LLLPSLSLLFLILKRRNRKTRFNLPRPGSGMPLFGETIGYLYKPYATTATLGD 66  
 Db 11 LLSVSTFLFATIFILLAGIARRKRRAPRLRPGSGMPLFGETIGYLYKPYATTATLGD 70  
 QY 67 FMOCHVSK-----YKIRSNLFGEPITVSADAGLNRFIL 101  
 Db 71 FVEKQIKFVSLCSVLLILKRPDNGFNEIRYGRIFSCSLFGMAVVASDAPDFNRFIM 130  
 QY 102 ONEGRLEPCSPRISGILGKMSMVLVIGDMHRDMSISLNLFLSHARLTILKDVERRHT 161  
 Db 131 ONEGRLEPCSPRISGILGKMSMVLVIGDMHRDMSISLNLFLSHARLTILKDVERRHT 161  
 QY 162 LFLVDSMOQONSIFSAODEAKKFTENLMAKHIMSMDPGEETEQKKEVYFMKGVASAPL 221  
 Db 191 LQTLISNFKDGEVLLQDICKRVAIHLVNWQLGSV--SESEVDMSQLESFVDGCLSVPI 249  
 QY 222 NLPGTAHYKALQSRAATILKTERKMEERKLDIKEDEDEEVEKTEDEAKMSDHYVRKOR 281  
 Db 250 DLPGFTYKAKKARKETIRKINKTIEKRLQKASD-----T 286  
 QY 282 TDDLLGVTLKHSNLSTEQIIDLISLFLAGHETSSVAIALAIFLQACPKAVEELREEH 341  
 Db 287 AGNYGLKLEESLPNDSMDFTIINLFGNETISTKMLFAVFLTNCPRAIMQOLLEEH 346  
 QY 342 LEIARAKKEGESELNMDYKKMDFTCVINETLRLGNVVFRLHKKALKDVRKYGYDIPS 401  
 Db 347 -----DRLAGMLTQMDYKTMDFTCVIDETLRGLGIALMREAKEDVSXYDYVPIK 399  
 QY 402 GWKVLPIYSAVHLDNSRYDQPNLFNPMRMO-----QONNGASSSGSFGSTWGNVMPFG 456  
 Db 400 GCFVVPPLSAVHLDNSRYDQPNLFNPMRMO-----FYCPFG 449  
 QY 457 GSPRLCAGSELALEMVFIHLVLFKFMNELAEDDOPFAFPVDPNGLPIRVSR 511  
 Db 450 GSTRFCAGSELALEMVFIHLVLFKFMNELAEDDOPFAFPVDPNGLPIRVSR 504



```

RESULT 4
D85429
cytochrome P450 like protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 02-Apr-2001
C:Accession: D85429
R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring
Nature 402, 769-777, 1999
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: A85001; MUID:20083488
A:Accession: D85429
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-457 <STO>
A:Cross-references: GB:NC_001268; NID:g2720586; PIDN:CAB80304.1; GSPDB:GN00140
C:Genetics:
A:Gene: AT4g36380
A:Map position: 4
C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology
C:Keywords: heme; iron; metalloprotein
F:396/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 32.1%; Score 860; DB 2; Length 457;
Best Local Similarity 36.8%; Pred. No. 2.6e-48;
Matches 179; Conservative 93; Mismatches 160; Indels 54; Gaps 8;

OY 39 LPPGSGPFLGETIGYL-----KPYATTLGDFMOOHVSKYKITYSNLFGEPTIVSA 92
D 1 PPGSGPFLGETIGYL-----KPYATTLGDFMOOHVSKYKITYSNLFGEPTIVSA 92
OY 93 DAGLRFLQDNGRLFECSYSPRSIGILGKMSMLVLDGMDHDMDSISLNFISHARLTI 152
D 2 PPGSGPFLGETIGYL-----KPYATTLGDFMOOHVSKYKITYSNLFGEPTIVSA 92
OY 56 DAENKAVLVQNHGNTFVAIPKSTIELGENSELISNGPHOKRLHTLIGAFURSPHLDR 115
D 1 PPGSGPFLGETIGYL-----KPYATTLGDFMOOHVSKYKITYSNLFGEPTIVSA 92
OY 153 LKQVERHTLEVLDSMOONSIFSAODEAKKFTFNLMARKHNSMDGEBETEOLKREYTF 212
D 1 PPGSGPFLGETIGYL-----KPYATTLGDFMOOHVSKYKITYSNLFGEPTIVSA 92
OY 116 ITRDIEASVLTILASMAQLPLVHGDDEIKMTFELVVKVSTSG-BDMNLKLEFEF 174
D 1 PPGSGPFLGETIGYL-----KPYATTLGDFMOOHVSKYKITYSNLFGEPTIVSA 92
OY 213 MKGVSAALNLPRTAYHAKALOSRATILKFERKMEERKLDIKEEDQEEBEVKT-----E 266
D 1 PPGSGPFLGETIGYL-----KPYATTLGDFMOOHVSKYKITYSNLFGEPTIVSA 92
OY 175 IKGILCIPIKPFGRFLYKSLAKERLIKMKVKKVVERQVAMTTSPADVDVLLRSGD 234
D 1 PPGSGPFLGETIGYL-----KPYATTLGDFMOOHVSKYKITYSNLFGEPTIVSA 92
OY 267 DEAEKSDHVRKORTDDDLGWLAKHNSLSTEQLDLISLFGHETSSAIALATFF 326
D 1 PPGSGPFLGETIGYL-----KPYATTLGDFMOOHVSKYKITYSNLFGEPTIVSA 92
OY 235 SEKQSPDEYSGK-----IVEMMLPGEETMPTAMTAVKE 270
D 1 PPGSGPFLGETIGYL-----KPYATTLGDFMOOHVSKYKITYSNLFGEPTIVSA 92
OY 327 LQACRAVEELREHLEIARAKKELESELNMDYKKMDFTQCVINETLRLGNVRFYLR 386
D 1 PPGSGPFLGETIGYL-----KPYATTLGDFMOOHVSKYKITYSNLFGEPTIVSA 92
OY 271 LSDNVALAKLVEENMEKRRRLTGE-EYKWTQVMSLFTQNVINETLRMANIINGVWR 329
D 1 PPGSGPFLGETIGYL-----KPYATTLGDFMOOHVSKYKITYSNLFGEPTIVSA 92
OY 387 KALKDVRKKGIDIPSGKVLVPLVISAHLNDRYDQNLFPNWRMOOQONGASSSGSFS 446
D 1 PPGSGPFLGETIGYL-----KPYATTLGDFMOOHVSKYKITYSNLFGEPTIVSA 92
OY 330 KALKDVEIKGILIPKGMVLASFISVHDEIDYDNPYOFDPWRMRDRINGSANSSIC--- 385
D 1 PPGSGPFLGETIGYL-----KPYATTLGDFMOOHVSKYKITYSNLFGEPTIVSA 92
OY 447 TWGNMYMFGGPRCACSELAKEMAVFIHHLVLFKFWELAEEDQPAFPVDPNGLP 506
D 1 PPGSGPFLGETIGYL-----KPYATTLGDFMOOHVSKYKITYSNLFGEPTIVSA 92
OY 386 -----FTPFGGGQRLCPLELUSKETISFLHLVLRYSM-TAEDEIVSFPVKKMRRLP 439
D 1 PPGSGPFLGETIGYL-----KPYATTLGDFMOOHVSKYKITYSNLFGEPTIVSA 92
OY 507 IRVSR 512
D 1 PPGSGPFLGETIGYL-----KPYATTLGDFMOOHVSKYKITYSNLFGEPTIVSA 92
OY 440 IRVATV 445
D 1 PPGSGPFLGETIGYL-----KPYATTLGDFMOOHVSKYKITYSNLFGEPTIVSA 92

RESULT 5
T48613
hypothetical protein F18022.190 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T48613
R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Le
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z24493

```

```

A:Accession: T48613
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-382 <BEV>
A:Cross-references: EMBL:DB
A:Experimental source: cultivar Columbia; BAC clone F18022
C:Genetics:
A:Map position: 5
A:Introns: 31/2; 139/3; 204/1; 224/3; 253/3; 280/1; 315/3; 349/2
A:Note: F18022.190

Query Match 30.4%; Score 815; DB 2; Length 382;
Best Local Similarity 34.2%; Pred. No. 1.7e-45;
Matches 159; Conservative 87; Mismatches 127; Indels 92; Gaps 4;

OY 45 GMPFLGETIGYLKPYATTLGDFMOOHVSKYKITYSNLFGEPTIVSADAGINRETLONE 104
D 2 GMPFLGETISFFKPHRSOSIGFLQQRVSRGKVKFNSICGKAVASCDQDELMTMIGNE 61
OY 105 GRLEPCSTPRISIGILGKWSMLVLDGMDHDMDSISLNFISHARLTIILLDVERHTLFV 164
D 62 GRLEPCSTPRISIGILGKWSMLVLDGMDHDMDSISLNFISHARLTIILLDVERHTLFV 164
OY 165 LPSMOONSIFSAODEAKKFTFNLMARKHNSMDGEBETEOLKREYTFEMKGVSAALNP 224
D 122 LKSMNCEVEERHKEVKTFTLSVMYNOLSTIKPEDPARLYVQDLSTYKGFISLPPLP 181
OY 225 GTAYHAKALOSRATILKFERKMEERKLDIKEEDQEEBEVKTDEAEKSDHVRKORTDD 284
D 182 GTGYTNAIKVRSNRNIHQNAIEDNNNAIREDPFLDSIISND----- 224
OY 285 DLGWLAKHNSLSTEQLDLISLFGHETSSAIALATFFLQACRAVEELREHLEI 344
D 225 -----EENHAI 230
OY 345 ARAKKESELNMDYKKMDFTQCVINETLRLGNVRFYLRHAKALDVRKIDIPSGMK 404
D 231 RAKKGDEL-LNMDYKKMDFTQCVINETLRLGNVRFYLRHAKALDVRKIDIPSGMK 288
OY 405 VLPVISAHLNDRYDQNLFPNWRMOOQONGASSSGSFSFTWGNMYMFGGPRLCAG 464
D 289 VEPITFAVHLDPSSLNENFEFPMRMTKT-----AEGGVAVCPG 329
OY 465 SELAKLEMAVFIHHLVLFKFWELAEEDQPAFPVDPNGLDIRV 509
D 330 GEIGKIQIAIFLHLVLSYRWKIKSDEMPIAPVVEFRGMLEI 374

RESULT 6
T07859
cytochrome P450 homolog - tomato
C:Species: Lycopersicon esculentum (tomato)
C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 16-Feb-2001
C:Accession: T07859
R:Bishop, G.J.; Harrison, K.; Jones, J.D.
Plant Cell 8, 959-969, 1996
A:Title: The tomato Dwarf gene isolated by heterologous transposon tagging encodes th
A:Reference number: Z16181; MUID:96266705
A:Accession: T07859
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-464 <BIS>
A:Cross-references: EMBL:U54770; NID:g1421740; PIDN:AA81070.1; PID:g1421741
A:Experimental source: strain GCR758
C:Genetics:
A:Gene: dwarf
C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology
F:273-436/Domain: cytochrome P450 homology <P45>

Query Match 28.4%; Score 761.5; DB 2; Length 464;
Best Local Similarity 33.4%; Pred. No. 6.2e-42;

```

Matches 170; Conservative 90; Mismatches 188; Indels 61; Gaps 10;

12 LLLPSSLLPFLILKRRN--RKRFRNLPKSGMPFLGTTGKLPYATITLG-DFM 68  
 5 LIFLSSFGCLIFCALRNMVKNQKNLPDGTGMPLFETETFLK-----LGPSFM 58  
 QY 69 QOHVSKYKIRYSLNFGPPTIVSADAGLNRFILONEGRFLFECSPRSIGLIGKMSMLV 128  
 Db 59 KNQARVGSFPRSHLIGCPITVSMDELNRYILVNEAGLVPGYQSMIDILGKNINAAV 118  
 QY 129 VGDHMRDRSISLNLSHARLTILKDYERHTLFVLDWMQONSIFSAODEAKKFTENLM 188  
 Db 119 NGSARKYRKGALLSLISPTMIRDLPLPDEMRSHLNM--DNKVYIDQEKTKMKAFSS 177  
 QY 189 AKHISMDPGEET---EQLKKEYTEPMKGVASAPLNLPGTAYHKALOSRAITLKFIERK 245  
 Db 178 LKQI-----AGLESTSLAOEFMSEFNVLVGLTSLPILPNTNHYHGFQARKIYNLLRTL 233  
 QY 246 MEERKLDIKEDEEVEEYKTEDEAEMSKSDHVRRKORTDDLLGWLKHS---NLSTEOI 301  
 Db 234 IEER-----ASKEIGHMDLGYLMNEEATRFKLTDEM 266  
 QY 302 LDILISLFPAGHETSSVAIALAIFLQACPAVEELREHLEIARAKKEGSELNMDY 361  
 Db 267 IDLITTIYSGEYETSTSMAYKYLHDHPVLELRKREHMAIREKKRP--EDPIDYNDY 324  
 QY 362 KKMDETQCVINETLGNVVRFLHRKALKDVRKYDIPSGMKVLPVISAHLNDSRYDQ 421  
 Db 335 RSMRTRAVILETSLATIVGVLKRTQDMEINCYIIPKGRIVYRELINADRLYPD 384  
 QY 422 PNLFPNPMQMOONNGASSSGSFSFTMGNNYMPFGGPRCLAGSELAKLEMAVFIHNLVL 481  
 Db 385 PYSFPMWMMKDS-----LEHONSFLVEGGGRQCPGKRLGVAELSTFLHYVT 433  
 QY 482 KFNWELADDQPFAPFPYDFPNGLPTRVS 510  
 Db 434 KYRWEITGDKLMPRVEAPNGLRIVS 462

RESULT 7

T04444  
 cytochrome P450 - Arabidopsis thaliana  
 N:Alternate names: protein T18B16.200; protein T5K18.10  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 28-Jul-2000  
 A:Accession: T04444; T05806  
 R:Bayan, M.; Benes, V.; Rechmann, S.; Borkova, D.; Ansoorge, W.; Bancroft, I.; Mewes, H.-W.  
 submitted to the Protein Sequence Database, April 1998  
 A:Reference number: Z15359  
 A:Accession: T04444  
 A:Molecule type: DNA  
 A:Residues: 1-457 <BEV>  
 A:Cross-references: EMBL:AL021687  
 A:Experimental source: cultivar Columbia; BAC clone T18B16  
 R:Bayan, M.; Van Der Schueren, J.; Chuang, Y.-J.; Voet, M.; Robben, J.; Voickaert, G.; Ba  
 submitted to the Protein Sequence Database, April 1998  
 A:Reference number: Z15453  
 A:Accession: T05806  
 A:Molecule type: DNA  
 A:Residues: 131-457 <BEV>  
 A:Cross-references: EMBL:AL022580  
 A:Experimental source: cultivar Columbia; BAC clone T5K18  
 C:Genetics:  
 A:Map position: 4  
 A:Insertions: 67/2; 173/3; 302/3; 358/1; 393/3  
 A:Note: T18B16.200; T5K18.10  
 C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology  
 C:Keywords: chromoprotein; electron transfer; heme; iron; monooxygenase; oxidoreductase  
 F:272-433/Domain: cytochrome P450 homology <P45>

Query Match 23.6%; Score 632; DB 2; Length 457;

Best Local Similarity 31.8%; Pred. No. 1.5e-33;  
 Matches 151; Conservative 92; Mismatches 176; Indels 56; Gaps 12;

QY 13 LLLPSSLLPFLILKRRNKRFRNLPKSGMPFLGTTGKLPYATITLGDFM 72  
 Db 10 LPAQSLFYLFRCLISORFSSKILPLPCTGMPEYVET---FQLYSQDP--NVFQSKQ 65  
 QY 73 SKYKIRYSLNFGPPTIVSADAGLNRFILONEGRFLFECSPRSIGLIGKMSMLVLD 132  
 Db 66 KRYSVFETHVLGCPVWISSPEAKFVLYTKSHLFTFPASKERMKGKQAFPHQGDY 125  
 QY 133 HRDMRSISLNLSHARLTILKDYERHTLFVLDWMQONSIFSAODEAKKFTENLM 192  
 Db 126 HAKLRKLVRAFMPESIRN--WVPIESIAQSLRSM--EGTIMINTYQEKTYTFVALLSI 183  
 QY 193 MSMDPGEET---EQLKKEYTEPMKGVASAPLNLPGTAYHKALOSRAITLKFIERK 249  
 Db 184 F---GKDEVLYRDLKRCYIILEKGYSMFVNLPGLTFHKSMAKARKLSOILARILSER 239  
 QY 250 KLDIKEDEEVEEYKTEDEAEMSKSDHVRRKORTDDLLG--WLKHSNLSTEOIDLILSL 308  
 Db 240 R-----QNGSSH-----NDLGSFMGKKEELTDQIDNITIGV 272  
 QY 309 LPAGHETSSVAIALAIFLQACPAVEELREHLEIARAKKEGSELNMDYKKMPTQ 368  
 Db 273 IFAARDTASVSMNLKYLAEINPVLEAVTEQMAI--RKDEEGES--LTWGDTRKMLPTS 330  
 QY 369 CVINETLGNVVRFLHRKALKDVRKYDIPSGMKVLPVISAHLNDSRYDQPFNPM 428  
 Db 331 RVIOETTLVAVSLSTFPEAEVDEYEGYLIIPKGVKYLPLRNIHSDIFSNPKRFPDS 390  
 QY 429 RMQOONNGASSSGSFSFTMGNNYMPFGGPRCLAGSELAKLEMAVFIHNLVLKF 483  
 Db 391 RFE-----VAPKRNTPFPGNGTHSCPGNELALEMSIMIHLLTTRY 432

RESULT 8

T02739  
 probable cytochrome P450 At2g29090 [Imported] - Arabidopsis thaliana  
 N:Alternate names: cytochrome P450 homolog T914.17  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 16-Feb-2001  
 A:Accession: T02739; D84692  
 R:Rounsley, S.D.; Lin, X.; Kaul, S.; Shea, T.P.; Fujii, C.Y.; Mason, T.M.; Shen, M.;  
 submitted to the EMBL data library, August 1998  
 A:Description: Arabidopsis thaliana chromosome II BAC T914 genomic sequence.  
 A:Reference number: Z14710  
 A:Accession: T02739  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-482 <ROU>  
 A:Cross-references: EMBL:AC005315; NID:93461834; PID:93461849  
 A:Experimental source: cultivar Columbia  
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Unayam, L.; Tallon,  
 euss, D.; Niemman, W.C.; White, O.; Elsen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
 Nature 402, 761-768, 1999  
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A:Reference number: AB4420; MUID:20083487  
 A:Accession: D84692  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-482 <STO>  
 A:Cross-references: GB:AE002093; NID:93461849; PID:AMC33235.1; GSPDB:GN00139  
 C:Genetics:  
 A:Gene: At2g29090; T914.17  
 A:Map position: 2  
 A:Insertions: 80/2; 187/3; 238/3; 321/3; 351/3; 413/3; 453/2  
 C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology  
 F:291-453/Domain: cytochrome P450 homology <P45>

Query Match 22.8%; Score 611; DB 2; Length 482;

Best Local Similarity 29.8%; Pred. No. 3.5e-32;  
Matches 153; Conservative 101; Mismatches 192; Indels 68; Gaps 13;

QY 16 PSLISLLFLILK-----RRNRKRNLPKSGMPFLGETIGYLPATTLGDMQ 69  
D 20 PALLTIVVVVVVLLKRWMLHMKOERLPPGSMGLPIYGET---RLTYEND-NSFFA 75  
QY 70 QHVKYKRYKRNLFGEPTIVSADAGLNRFILONEGRLEFCSYPRSIGLIGKMSMLV 129  
D 76 TRONKYGDIETHILGCVWISSPEARAVLSKAHLFPYPSKERMIGPEALFPHQ 135  
QY 130 GDMRDM-RSISLNFSLHARLTLKDVNRHTFLVDSMOONSIFSAODEAKKFTENIM 188  
D 136 GPVSTLKLRLVQSSFMPSALRPV--SHIELLVLOTLSMSWSQKSINLEVMKRYAFV 193  
QY 189 AKHMSMDPGEE---TEOLKKEYVTMKGVASAPLPGTAHYKALOSRATITIKFERK 245  
D 194 ---TMSAFGDEEPTTIDVILKLYORLEGYNSMPLDPLGTLFHKSMKARIELSELK 250  
QY 246 MEERKLDKEEDDEEVEKTEDEAEMSKSDHVRKQRTDDLLGWL-----KHSNLSTEQ 300  
D 251 IEKRENGREE-----GGLGVLLAKKDKRNGLSQ 283  
QY 301 ILDLILSLFAGHETSSVAIALAIFLOACPRAVEELREHLEI-ARAKKEGESELNMD 359  
D 284 IADNIGVFAATDTTASVLTWMLKYLHDHNLQEVSRQEFSTIRKTKKE--NRISME 341  
QY 360 DYKMDFTQCYINETLRGNVRFLLKRAKADVKRYGVDIPSGMKVLPVISAHLDSRY 419  
D 342 DTRKMPLTRVIOETLRKASVLSFTFRVADVEDYGLIKRGMKVLPLFRHHSSEF 401  
QY 420 DQPLNFPMWMOQONNGSSGFSFGWNNYMPFGGPRLLCAGSELAKTEMAVFIHHL 479  
D 402 PDPEKFDSDRE-----VAPKPYTMEFGVHSCGSELAKTEMLLHL 448  
QY 480 VLKFNWELAEDDQFAF-PFVDFPNGLPYRSRI 512  
D 449 TTSFRWEYIGDEGIQYGFPPVKKGLPIRYTPI 482

# RESULT 9

C71417  
cytochrome P450 d13695c - Arabidopsis thaliana  
N:Contains: oxidoreductase (EC 1.-.-.-)  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
A:Variety: Columbia  
C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000  
C:Accession: C71417  
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Berokamp, R.; Dirk  
P.; Medler, H.; Medler, E.; Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terry, N.; Giel  
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaefer, M.; Funk, B.  
Nature 391, 485-488, 1998  
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Pulidomench  
erhoff, A.; Moore, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans  
C.; Chaiwazis, N.  
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal  
A:Reference number: A71400; MUID:86121113  
A:Accession: C71417  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-487 <BEV>  
A:Cross-references: GB:297338; NID:g2244870; PIDN:CAB10309.1; PID:g2244888  
C:Genetics:  
A:Gene: d13695c  
A:Map position: 4COP9-4G3845  
C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology  
C:Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase  
F:433/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 21.7%; Score 581; DB 1; Length 487;  
Best Local Similarity 27.3%; Pred. No. 3.1e-30;  
Matches 143; Conservative 100; Mismatches 205; Indels 76; Gaps 11;

QY 17 SLISLLFLILKRRNRKRNLPKSGMPFLGETIGYLPATTLGDMQHYSKY 76  
D 3 SLFLKTFHWWYQWNPENKLPKSGMPFLGETIGYLPATTLGDMQHYSKY 62  
QY 77 -----KIRSNLFGEPTIVSADAGLNRFILONEGRLEFCSYPRSIGI-----LG 121  
D 63 ADFSINHSFRTSLFGKALISMDMLN-----LEMAKANSYPGYKVIKRLFG 112  
QY 122 KWSMLVGVDMHMRDRSISLNFSLHARLTLKDVNRHTFLVDSMOONSIFSAODEAK 181  
D 113 ENNFLOSKESSHKVRNLPOLLGPOGLSKSMIEDVDLARTYMEGARNGYLDVKETSS 172  
QY 182 KFTFNMAKHLIM-SMDPEEETEQKKEYVFMKGVASAPLPGTAHYKAL-----232  
D 173 KILGCLCKKVMGEWP--EAKELALCWRYFQSGWFRFPLNLPSTGYKKMKKVLVQYT 230  
QY 233 -----QSRATILKFERMEERKLDKEEDDEEVEVTEDEAEMSKSDHVRKQRTDDL 286  
D 231 EADISQARKKMKLRTVLTKRASGELEGFNIIIGEMGE-----274  
QY 287 LGWIKHSNLTETQTLILSLFAGHETSSVAIALAIFLOACPRAVEELREHLEIAR 346  
D 275 -----GFTMSVENAVEYITFFLVANETTPRIILATVYKFTSDHPKVKQELQREHEIAR 328  
QY 347 AKKEGESELNMDQYKMDPQCYINETLRGNVRFLLKRAKADVKRYGVDIPSGMKVLP 406  
D 329 GKAE-KEGGLTWEDKSNHFTQVINESLRITSTAPVLRLEHFOGDTYIPGWFPM 387  
QY 407 PVISAVHDSNRDQPLNFPMWMOQONNGSSGFSFGWNNYMPFGGPRLLCAGSEL 466  
D 388 G-YPIHHSKKEDEPYAFNFMWMEGKDLGALV-----KFTIFGAGRILCVAE 437  
QY 467 LAKEMAVFIHHLVLKFNWELAEDDQFAFPVDFPNGLPYRSRI 510  
D 438 FAKQMAVFIHHL-FRYRMSKSGTTRISFPLMPGCGDVOIS 480

# RESULT 10

A84859  
probable cytochrome P450 (imported) - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 23-Mar-2001  
C:Accession: A84859  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y  
M.; Koo, H.; Mofat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon,  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487  
A:Accession: A84859  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-485 <STO>  
A:Cross-references: GB:A8002093; NID:g4512670; PIDN:AAD21724.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g42850  
A:Map position: 2  
C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology  
C:Keywords: heme; iron; metalloprotein  
F:432/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 21.6%; Score 579.5; DB 2; Length 485;  
Best Local Similarity 28.8%; Pred. No. 3.9e-30;  
Matches 147; Conservative 107; Mismatches 203; Indels 53; Gaps 11;  
QY 12 LLLPSLLSLFL-----ILKRRNRKRNLPKSGMPFLGETIGYLPATTL- 64  
D 14 LCIAATISSTLEFFRKKNHRETKIKQKKR--LLPGEMGLPWIGETDYPKAQKSNRVF 71  
QY 65 GDFMOQHSXKGYKIRSNLFGEPTIVSADAGLNRFILONEGRLEFCSYPRSIGLIGKWS 124  
D 65 GDFMOQHSXKGYKIRSNLFGEPTIVSADAGLNRFILONEGRLEFCSYPRSIGLIGKWS 124

Db 72 EDFVNPRIKHNIFKTRIMSGPTIVNGAEANRLILSNEEFLVSSWPSSVOLMGNC 131  
 QY 125 MLVYGDHNRMRSLINFLSHARLITLLKDVENHTEFLVDSMOQNSIFSAODEAKFT 184  
 Db 132 IMAKGEGHARVLRGLVANSLSYIGLESILPKLDVTKFHNHETEMGKEEISLYRSKAYLT 191  
 QY 185 ENLAKNHI--MSMDGEBETEOLKKEYVTFMKGVYSAPLPGTAYHAKLSRATILKEI 242  
 Db 192 FTVVECELYGKVELGMLEY-----FERYLEGVAFALVERPCSFARAKKARLEIEFLL 245  
 QY 243 ERKMEERLIDKEDEQEEVEKTEDEAMSKSDHVRKORTDDLLGVNLKHSNJSTEOIL 302  
 Db 246 VGKREKREKREKGEAE-----KPNITLFSRLVEELIKGV-----ITEEEVV 287  
 QY 303 DLILSLFAGHETSVAIALAIFLQACPKAVEELREHELEIARAKKELGSE--LMDY 361  
 Db 288 DNMVLYVAADHTTSYANSMTFKMLAOHPTCRDILLOEHAOI---KANGGCEITLYVEDV 344  
 QY 362 KKMDFOCVINETLRGLNVVRFHLHRKALDVRKGYDIPSGMKVLPVISAHLNDSRYDQ 421  
 Db 345 KKMYSMQVAVRETMRLSPRIGSPFRKAVADIDGYTTPKGMKILMTYGHVNPETIFOD 404  
 QY 422 PNLNPRHMQOONNGASSSGSGSEFTMGNNMFRGGGRLCAGSELAKLEMAVFIHLVL 481  
 Db 405 PMSDFPRFEDKPIQAYT-----YLPFGGGRLCAGHQLAKISILVFMHFEVYT 451  
 QY 482 KFNNELEDDOPFAFPEVDFPN--GLPIRVS 510  
 Db 452 GFDMSLVYPRDETISMDPLPFSLCMPIKIS 481

## RESULT 11

H86185  
 hypothetical protein [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: H86185  
 R:Theo1018, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Hansen, N.F.; Hughes, B.; Huitzer, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; Li, J.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719  
 A:Accession: H86185  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-490 <STO>  
 A:Cross-references: GB:AE005172; NID:g238581; PIDN:AAB71462.1; GSPDB:GN00141  
 C:Genetics:  
 A:Map position: 1  
 C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology

Query Match 21.6%; Score 579.5; DB 2; Length 490;  
 Best Local Similarity 28.6%; Pred. No. 4e-30;  
 Matches 133; Conservative 107; Mismatches 192; Indels 33; Gaps 9;  
 QY 34 KTRFNLPRGKSGMPFLGETTIGYLKPYATATLGDPMQOHVSKYGR--LYRSNLGFEPTIVS 91  
 Db 40 ENRHYLPRGDLGMPFGICMLSLFAKTSDDPSFTRLIKRYGPKGITYKAMFENPSITIV 99  
 QY 92 ADAGLNRFILONEGRLEECSTPRISIGTILGKWSMLVGVDMHRMRSISLFLSHARLT 151  
 Db 100 TTSOTCRVRLTDD--AFKPGWPTSTMELIGKRSFVGSFEHKKRLRLTLARVNGHEALS 158  
 QY 152 ILKLDVERHTEFLVDSMOQNSIFSAODEAKFTFNLAKNHMSMDPGEELTEOLKKEYVT 211  
 Db 159 TYPIYLENVTIADKMTKMGEEFLLHRLKLTERRI--MYTIFLSESEENVADALREYTA 217

QY 212 FMKGVSAPLNPGTAYHAKLSRATILKTERKMEERKLDIKEDEQEEVEKTEDEAEM 271  
 Db 218 LNYGRMAAVNIPGFARALKARLTLVAFOSTYTER-----NOKKNILSKKMDL 271  
 QY 272 SKSDHVRKORTDDLLGVNLKHSNJSTEOILDLISLTFAGHETSVAIALAIFLQACP 331  
 Db 272 DNLNVR-----DEK-----GKTLDEEITIDVLMYLNAGHESGHTIMATVFLQENR 320  
 QY 332 KAVEELREHELEIARAKKELGSEELMDDYKAMDFOCVINETLRGLNVVRFHLHRKALD 391  
 Db 321 EYLORAKAEQEMIIKSRPE--GQKLSLKEKTRKMEFLSQVDETLRVITFSLTAFREACTD 379  
 QY 392 VRYGYDIPSGMKVLPVISAHLNDSRYDQNLNPRHMQOONNGASSSGSGSPSTMGNC 451  
 Db 380 VEMNGYILPRKGMKYLTVRVDYHIDEPVPRPKRDPARW---DNG-----FVPRAGA 428  
 QY 452 YMPFGGRLCAGSELAKLEMAVFIHLVLKFNNELEDDOPFAF 496  
 Db 429 FLPRGAGSHLCRGNDLAKLEISIFLHFLKLYQVRSNPECPVNY 473

## RESULT 12

cytochrome P450-like protein - Arabidopsis thaliana  
 N:Alternate names: protein F14D17.40  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 18-Aug-2000  
 C:Accession: T48973  
 R:Jordan, N.; Bangert, S.; Wiedemann, R.; Voss, H.; Unseld, M.; Mewes, H.W.; Rudd, S.  
 submitted to the Protein Sequence Database, April 2000  
 A:Reference number: Z25008  
 A:Accession: T48973  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-455 <JOR>  
 A:Cross-references: EMBL:AJ353992; GSPDB:GN00061; ATSP:F14D17.40  
 A:Experimental source: cultivar Columbia; BAC clone F14D17  
 C:Genetics:  
 A:Gene: ATSP:F14D17.40  
 A:Map position: 3  
 A:Insertions: 70/2; 178/3; 229/3; 312/3; 343/3; 346/1; 381/3  
 C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology

Query Match 21.5%; Score 577; DB 2; Length 455;  
 Best Local Similarity 28.9%; Pred. No. 5.2e-30;  
 Matches 145; Conservative 84; Mismatches 173; Indels 100; Gaps 10;

QY 32 NRKTRFNLPRGKSGMPFLGETTIGYLKPYATATLGDPMQOHVSKYKTYRSNLGFEPTIVS 91  
 Db 28 NPKNSNGKLPSPSGMGPFLIGETLDFPKRPGFYEISPYLKKMLRKGPLEFRTNIIAGVTVVS 87  
 QY 92 ADAGLNRFILONEGRLEECSTPRISIGTILGKWSMLVGVDMHRMRSISLFLSHARLT 151  
 Db 88 TDXDVNMELIRQENKSPFLSYDPGLMKPRKDSFLKGINHKKIKQITLLHLLSEGLK 147  
 QY 152 ILKLDVERHTEFLVDSMOQNSIFSAODEAKFTFNLAKNHMSMDPGEELTEOLKKEYVT 211  
 Db 148 KILKMDVTRTRELHLSKAKTGRLDVKNVSKLITAHILPRKMS-----NLKPGQA 198  
 QY 212 FMKGV-----VSAPLNPGTAYHAKLSRATILK--FIERKMEERKLDIKE 255  
 Db 199 KLMDIFAAFTFDWERTSYLSAGKGLVNTLM--ACRGMREIKIYIMRKRSSEKY----- 252  
 QY 256 EDOEEVEKTEDEAEMSKSDHVRKORTDDLLGVNLKHSN-----LSTEOILDLISLFL 310  
 Db 253 -----DDFLNPAIEESEKAGELNMDNAILTLFLTSC 284  
 QY 311 AGHTSSVAIALAIFLQACPKAVEELREHELEIARAKKELGSEELMDDYK--KMDFOCV 369  
 Db 285 VTQDTSKAICLAVKFLLENPKVLAELKKEH--EVLESREDEKGGVWEEYRHKMFTTN- 342

QY 370 VINETLRLGNVYRLHRAKLDVRYKGYDIPSGMKVLPVISAHLNDSKYDQNLFPNPR 429  
 Db 343 -----MKSGETIPAGWIMTIPSVVHDPPELEPNPPEENPMP 379  
 QY 430 MOONNNGASSSGSFSFTWGNVMPFGGPRLCAGSELAKEMAVFIHHLVTKFNMWELAE 489  
 Db 380 WEGKELRAGS-----KTFMWFGTGLRQCAGAEFAFIQISVFLHLLVTYTNFSLHQ 429  
 QY 490 DDQPAFPFVDFPNCPLPIRVSR 511  
 Db 430 DCEVLRAVPAHLNPGNISINISK 451

## RESULT 13

B84733  
 Probable cytochrome P450 (imported) - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
 C:Accession: B84733  
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanhaken, S.E.; Umayam, L.; Tallon, L.;  
 Guss, D.; Niernan, W.C.; White, O.; Elsen, J.A.; Salzberg, S.L.; Fraser, C.M.; Ventner, J.  
 Nature 402, 761-768, 1999  
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A:Reference number: A84420; MUID:20083487  
 A:Accession: B84733  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-489 <STO>  
 A:Cross-references: GB:AE002093; NID:q3831452; PIDN:AAC69934.1; GSPDB:GN00139  
 C:Genetics:  
 A:Gene: At2g32440  
 A:Map position: 2  
 C:Superfamily: Synecocystis cytochrome P450 slr0574; cytochrome P450 homology

Query Match 20.4%; Score 547.5; DB 2; Length 489;  
 Best Local Similarity 27.6%; Pred. No. 4,6e-28;  
 Matches 146; Conservative 105; Mismatches 211; Indels 67; Gaps 14;

QY 8 TLPLPLLSLSLLFLI-LKRRN-----RKTRFNLPGKSGMPFLGETTGYLK 57  
 Db 4 TGLIIMWFP-LIILGLEVLKWLKRVNWIYKSGKEKKHYLPBGDLGWPVIGNMMSFLR 62  
 QY 58 PYTATTLDFMQOHKSKG--IYRSNLFGEPTIYSADAGLNFILQNGRLEFESYPS 115  
 Db 63 AKTSDPESEFISYTRRGRTGIRAHMGPCVLTPTPETCRVLTDD-AFHIGMPKS 121  
 QY 116 IGGILGKMSLVLVGDHMDRSISINFLSHARLRTILKDYERHTLVFLDSMOONSIFS 175  
 Db 122 TMKLGKRSFVGSFPEHRLRLRTSAPYNGPEALSVYQFTEBYNTDLLEKSKKGELE 181  
 QY 176 AODEAKKFTFNMAKHMSMDPEEETEQLKREYTEMKGVASAPLNLPGTAVYHKAQSR 235  
 Db 182 FLSHRKLFLFYI-MYIFLSESEHVMDSLEREYTNLNGVBRAMINLPGFAYHRLKAR 240  
 QY 236 ----ATILKFERKKMEERKLDI--KEEOEEREEVKTEDAEKSKSHVAKQRTDDDLGW 289  
 Db 241 KKLVAFOSIVNRRNRKONISSNRKMDLNLIDVKE-----NGRVLDD---- 286  
 QY 290 VLKHSNLSLEQIIDLILSLFAGHETSSVALAIFFLQACPKAVELREHLEIRAKK 349  
 Db 287 -----EELIIDLIMLNGHSSGHLTMTATILKQHEHMILOKAKEEDERT--VAK 335  
 QY 350 ELGSELDNDYDKKMDFTQCVINETLRGNVYRLHRAKLDVRYKGYDIPSGMKVLPV 409  
 Db 336 RAPGCKLTAKTERREMYLSQVIDETLRVTFSLTAFAREKASDVOMDGYIIPKGMKVLTF 395  
 QY 410 SAVHLDNSKYDQNLFPNPRMOONNNGASSSGSFSFTWGNVMPFGGPRLCAGSELAK 469  
 Db 396 RNWHLDPPEIYPPKPFDSRME-----GTPRAGTFLLPGCLSHLCPGNDLAK 443  
 QY 470 LEMAVFIHHLVLFNMWELAEDDQPAFPFVDFPNCPLP-----IRVSRL 513

Db 444 LEISFLHHLFLKRYVERSNPGCPVAF-----LPNHRKPNCLARITRTM 488

## RESULT 14

D96813  
 Hypothetical protein T30F21.17 (imported) - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: D96813  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,  
 ansen, N.F.; Hughes, B.; Hultzer, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Maiti, R.; Matzla  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Ventner, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719  
 A:Accession: D96813  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-460 <STO>  
 A:Cross-references: GB:AE005173; NID:q4836883; PIDN:AAD30586.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: T30F21.17  
 A:Map position: 1

Query Match 18.6%; Score 497.5; DB 2; Length 460;  
 Best Local Similarity 26.2%; Pred. No. 7,3e-25;  
 Matches 133; Conservative 102; Mismatches 205; Indels 67; Gaps 12;

QY 13 LPLPSLSLSLLFLILKRRNRKTRFNLPGKSGMPFLGETTIGLKPYTATTLDFMQOHV 72  
 Db 9 MLMLVAVVYRISHWLRLWMSNPKCPKLPFGSGMGPFLIGETLDFEKKCGVGIPTPKKM 68  
 QY 73 SKYGIYRSNLFGEPTIYSADAGLNFILQNGRLEFESYPSISIGIIGKMSLVLVGDM 132  
 Db 69 IRGPLEFRINIGSKTVVSTDPDVIIHQIFROENTSELGYPDIFVAVFGKDNILFKEVEI 128  
 QY 133 HRDMRSISNFLSHARLRTILKDYERHTLVFLDSMOONSIFSADQEAFFENLMAKH 192  
 Db 129 HRYLOKITMOLIGSEGLKQTMGLGMDKATRDHIRSIASOGSFRVREVENLVAVYTPRL 188  
 QY 193 MSMDPEEETEQLKREYTEMKGVASAPLNLPGTAVYHKAQSRATILKFERKMEERKLD 252  
 Db 189 ISNLRPEQSKLIDN-----LNATRLDMFKSF-LRLSTWKAVTAKSR----- 231  
 QY 253 IKEEDOEEREEVKTEDAEKSKSDHVRKQRTDDDLGWLKH-----SNLSTQIIDLILS 307  
 Db 232 -----EBAIQVMKQVLMKRRKTRKQ---EDFLNTLLELEKDSFPDQSAIHLPL 281  
 QY 308 LFLAGHETSSVALAIFFLQACPKAVELREHLEIAAKKELGSELNMDYK-KMPF 366  
 Db 282 LAFALREGTSCTALAVKFIKDPKVLAEKREKAIIVNRKD-KEAGVSWEERYRHNTPE 340  
 QY 367 TOCVINETLRGNVYRLHRAKLDVRYKGYDIPSGMKVLPVISAHLNDSKYDQNLFP 425  
 Db 341 TINVSNSEVRLANTPTPLLRKAVQDVEIKGY-----YLEVTHMGFG----- 382  
 QY 426 NPMRMOONNNGASSSGSFSFTWGN-NYMPFGGPRLCAGSELAKEMAVFIHHLVLFKN 484  
 Db 383 --WLMD-----GKEMTWGSKTFMAFGYVRLCVGAFFSRLOMALFIHHLVAVYD 429  
 QY 485 WELAEEDQPAFPFVDFPNCPLPIRVSR 511  
 Db 430 FSNVQDSSEITIRSPHOYTQDLINISQ 456

## RESULT 15

102263  
 C:cytochrome P450 DMARF3 - maize  
 N:Contains: oxidoreductase (EC 1.-.-.-)  
 C:Species: Zea mays (maize)  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000  
 C:Accession: T02263  
 R:Winkler, R.G.; Helentjaris, T.  
 Plant Cell 7, 1307-1317, 1995  
 A:Title: The maize dwarf3 gene encodes a cytochrome P450-mediated early step in gibberellin biosynthesis  
 A:Reference number: Z14648; MUID:96004534  
 A:Accession: T02263  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Residue type: mRNA  
 A:Molecule: 1-519 <MIN>  
 A:Cross-references: EMBL:U32579; NTD:9987266; PIDN:AAC49067.1; PID:9987267  
 A:Experimental source: strain B73  
 C:Genetics:  
 C:Gene: dwarf3  
 C:Function:  
 A:Description: Involved in an early step in gibberellin biosynthesis  
 A:Pathway: gibberellin biosynthesis  
 C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology  
 C:Keywords: oxidoreductase  
 I:325-468/Domain: cytochrome P450 homology <P45>  
 I:325-468/Domain: cytochrome P450 homology <P45>

Query Match	18.3%	Score 491.5	DB 1	Length 519
Best Local Similarity	28.2%	Pred. No 2.1e-24		
Matches 129	Conservative 91	Mismatches 189	Indels 49	Gaps 11
QY	36	RNNLPKSGWPELGETGYLKPYPATYATLLGDGFMQHVSKYG--IYRSNLFGEPTI-VSA	92	
	:	:	:	:
Db	69	RARLPGEWGPVLYGGMMAFLRAFSGKPRDAFIASEVFRFGCTGVYRSFMESPVLVLT	128	
	:	:	:	:
QY	93	DAGLRFLIONGGRFEGESYPRSIGIGKSMVLVVGOMHNDMSISLNFASHARLRI	152	
	:	:	:	:
Db	129	AEGCQVLLDDDA--FVYGWPKATYALVGPFSFVAMPDEHNRIRIKTLAPINFDALTG	186	
	:	:	:	:
QY	153	LKIDVERHTLEVLDWM--OONSIFSADDEAKKFTFMIAKHIMSMDPGEEETBOLKEVYT	211	
	:	:	:	:
Db	187	YLPFDRIYVTSLSRMAADHGSGVEFTELIRMTFKIIVY-IFLGADQATTPALRSYTE	245	
	:	:	:	:
QY	212	FMKGVSAPALPGVAHYHKALOSRAITLIKFERKKKEPRLDIKE-----EQGEEEVYT	265	
	:	:	:	:
Db	246	LNYGRMAAINTPGATYALGALRARRRLVAVLQGVLDERRARAKVSGGVDMMRLLEA	305	
	:	:	:	:
QY	266	EDAEAMSKSDHYRKQRTDDLLGWVLKHSNLTSTEOIDLTLISLFRAGHETSSVAIALAF	325	
	:	:	:	:
Db	306	QDE-----RGRHLDDD-----ELIDVLVWYLAAGHSSCHITMATV	342	
	:	:	:	:
QY	326	FLQACPKAVEELREHLLIARAKKKEGSELNMDYDKKMDFTQCYNITELRIGNVYREIH	385	
	:	:	:	:
Db	343	FLQENPDMFARAKAQEALIMRSIPS--SGRGLTLRFKRMEYLSQVYIDELRLVNTSFVSF	401	
	:	:	:	:
QY	386	RKALDVRKYKGDIPEGKAVLPVIAVILNDNSRYDQPLFNFMWQOONNGASSSGSGSF	445	
	:	:	:	:
Db	402	RGATRDVEVNGILPIKGKKVQLWYMSVHMDPQVYRDPPIKFDPSRNE-----CHSRAGTF	456	
	:	:	:	:
QY	446	STWGNMYNPPGCGPRLCAGSELAKLEMAVFIHHVLKRF	483	
	:	:	:	:
Db	457	-----LAFGIGARLCPGNDLAKLEIVFVLIHNFLLGY	487	
	:	:	:	:

Search completed: October 17, 2001, 22:08:02  
Job time: 8206 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2001, 22:08:06 ; Search time 30.88 Seconds

(Without alignments)  
569.076 Million cell updates/sec

Title: US-09-502-426-2

Perfect score: 2681  
Sequence: 1 MFETEHHTLLPLLLPLSLLS.....FAFPVDFPNGLPIRVSRLL 513

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	978.5	36.5	472	1	C901_ARATH
2	761.5	28.4	464	1	CP85_LYCES
3	579.5	21.6	490	1	C883_ARATH
4	503	18.8	492	1	CP26_BRARE
5	491.5	18.3	519	1	C881_MAIZE
6	461	17.2	497	1	CP26_MOUSE
7	438	16.3	497	1	CP26_HUMAN
8	324.5	12.1	520	1	YRV2_CAEEL
9	321.5	12.0	518	1	YRVA_CAEEL
10	316	11.8	500	1	C912_ARATH
11	312.5	11.7	504	1	CP32_RAT
12	312	11.6	501	1	CPJ5_MOUSE
13	310.5	11.6	520	1	YRV1_CAEEL
14	303	11.3	501	1	CPJ6_MOUSE
15	302	11.3	494	1	CPA8_MESAU
16	299.5	11.2	492	1	CPBC_RAT
17	297	11.1	491	1	CPB9_MOUSE
18	294.5	11.0	491	1	CPB4_RABIT
19	293	10.9	491	1	CPB1_RAT
20	291.5	10.9	530	1	CPJ1_YEAST
21	290.5	10.8	502	1	CPB3_ARATH
22	289	10.8	503	1	CPJ9_RAT
23	288.5	10.8	492	1	CPB1_MOUSE
24	288.5	10.8	503	1	CPJ3_HUMAN
25	288.5	10.8	538	1	CPJ8_DROME
26	288	10.7	459	1	CPXN_ANASP
27	287.5	10.7	520	1	YRV5_CAEEL
28	287	10.7	506	1	CPJ2_HUMAN
29	286.5	10.7	502	1	CPJ2_HUMAN
30	286.5	10.7	533	1	CPJ2_HUMAN
31	284.5	10.6	505	1	C762_SOLME
32	281.5	10.5	504	1	CPJ3_MOUSE
33	281	10.5	528	1	CPJ1_CANTR

34	280.5	10.5	504	1	CPJ3_MOUSE	064459 mus musculus
35	280	10.4	499	1	C771_SOLME	P37123 solanum mel
36	280	10.4	500	1	CPJ1_RABIT	P52786 oryctolagus
37	279.5	10.4	513	1	C773_SOYBN	048928 glycine max
38	279.5	10.4	519	1	YRV8_CAEEL	027520 caenorhabdi
39	279	10.4	501	1	CAD2_DROME	027589 drosophila
40	279	10.4	502	1	CPJ3_RAT	P51590 rattus norv
41	278.5	10.4	501	1	CPJ3_RABIT	P11707 oryctolagus
42	278	10.4	470	1	CPB4_MOUSE	062397 mus musculus
43	278	10.4	503	1	CPJ3_MOUSE	064464 mus musculus
44	277.5	10.4	491	1	CPB5_RABIT	P12789 oryctolagus
45	276	10.3	491	1	CPB2_RAT	P04167 rattus norv

## ALIGNMENTS

RESULT	ID	STANDARD	PRT	472 AA.
1	C901_ARATH			
AC	Q42569			
DT	15-DEC-1998 (Rel. 37, Created)			
DT	15-DEC-1998 (Rel. 37, Last sequence update)			
DT	15-DEC-1998 (Rel. 37, Last annotation update)			
DE	CYTCHROME P450 90A1 (EC 1.14.-.-)			
GN	CYP90A1 OR CYP90 OR CPD.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;			
OC	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;			
OC	Brassicales; Brassicaceae; Arabidopsis.			
OX	NCBI_TaxID=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV. COLUMBIA.			
RX	MEDLINE=96200769; PubMed=8612270;			
RA	Szekeeres M., Nemeth K., Koncz-Kaiman Z., Mathur J., Kauschmann A.,			
RA	Altman T., Redei G.P., Nagy F., Schell J., Koncz C.;			
RT	"Brassinosteroids rescue the deficiency of CYP90, a cytochrome P450,			
RT	controlling cell elongation and de-etiolation in Arabidopsis.";			
RL	Cell 85:171-182(1996).			
CC	-1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).			
CC	-----			
DR	EMBL; X87367; CA60793.1; -			
DR	EMBL; X87368; CA60794.1; -			
DR	InterPro; IPR001128; -			
DR	Pfam; PF00067; P450; 2.			
DR	PROSITE; PS00086; CYTOCHROME_P450; 1.			
KW	Oxidoreductase; Monooxygenase; Heme.			
FT	BINDING 418 418 HEME (BY SIMILARITY)			
SO	SEQUENCE 472 AA; 53785 MW; 41A73F46D64E343F CRC64;			
Query Match	36.5%; Score 978.5; DB 1; Length 472;			
Best Local Similarity	41.9%; Pred. No. 1.3e-55;			
Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;				
OY	12 LLLPLSLLLFLLKRRNRKTRFNLPPKSGMPFLGFTIGYLYKPYATATLGGFMQOH 71			
DB	7 LLLSSIAAGFLD--LLRRTRYRRMGLPGLIGLIGFTFOLIGAVKTEMPDFIDER 63			
OY	72 VSKYKTRINSNFGERTVVSADAGNRLTNEGSLFECSTPRISGGLGKWSMLVLYGD 131			
DB	64 VARYSVFTMFLFGERTVVSADPETNRFVLNKGKLFECSTPASICNLGKHSLLLMGMS 123			
OY	132 MHRDRSISLNFSLSHARLRTILKDVRRHTLFVLDSMOONSIFSAODBAKKFTFNLMAKH 191			

```

Db 124 LHKRHSITMSFANSIIKIDMLDIDLVFNDSMSSVLL--MEBAKKITFEIYKQ 181
Qy 192 IMSMDPGEEETEQAKKEVTPMKGVASAPLPLPGTAHYKALQASRATILKFERKMEBEKL 251
Db 182 LMSDPPG-EMSESILRKEVLVIEGFSLPLPLFSTYRKALQAR-----RKVAELT 232
Qy 252 DIKEDDQEEFEVKTEDEAEMSKSDHVRKQRTDDLLGVLSHNSLSTEQIIDLILSLFA 311
Db 233 VVWAKRREBEAGAE-----RKDMIALALADDSGSDDEIYDLVALLVA 278
Qy 312 GHETSSVALAIFLQACPAVEELREHELEIARAKKEGESEELNMDYKMDFTQCVI 371
Db 279 GYETSTITMLAVKFLFETPLALQKKEHEKIRAMKSD--SYLSEMGDYKSMPEFTQCVV 336
Qy 372 NETLRGLVAVFLRKALKADVRYKGYDIPSGMKVLPVSAVHLNDSRYDQNLNPNPMRQ 431
Db 337 NETLRVAVIIGGVERRAATDVEIKGYKIPKGMKVSSFRVAHDLPNHFKDARTNPNMRQ 396
Qy 432 QONNASSSGSGSFSTWGN--YMPFGGPRLCAGSELAKLEMAVEIHLVLKFMNELAED 490
Db 397 -----SNSVTTGGSNVFTFPGGSPRLCPGVELARVALSVFLHRLVYGFSSVPAEQ 446
Qy 491 DQPAFPEFVDPNGLPFRVSR 511
Db 447 DKLVFEPTTRQKRYPIFKR 467

```

```

RESULT 2
CP85_LYCES STANDARD: PRT: 464 AA.
ID CP85_LYCES
AC Q43147;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CYTOCHROME P450 85 (EC 1.14.-.-) (DWARF PROTEIN).
GN CYP85 OR D.
OS Lycopersicon esculentum (tomato).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
OC Solanales; Solanales; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. GCR758;
RX MEDLINE=96266705; PubMed=6672892;
RA Bishop G.J., Harrison K., Jones J.J.G.D.;
RT "The tomato Dwarf gene isolated by heterologous transposon tagging
encodes the first member of a new cytochrome P450 family.";
RL Plant Cell 8:959-969(1996).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U54770; AAB17070.1; -
CC InterPro: IPR001128; -
CC Pfam: PF00067; P450.1.
CC PRINTS: PR00385; P450.
CC PROSITE: PS00086; CYTOCHROME_P450.1.
CC OXidoreductase; Monooxygenase; Membrane; Heme.
CC BINDING 414 414 HEME (BY SIMILARITY).
CC SEQUENCE 464 AA; 53706 MW; D2B21AAB7B14E94 CRC64;

```

Query Match 28.4%; Score 761.5; DB 1; Length 464;  
 Best Local Similarity 33.4%; Pred. No. 9.5e-42;  
 Matches 170; Conservative 90; Mismatches 188; Indels 61; Gaps 10;

```

Qy 12 LLLPLSLLLFLILLKRN--RKTRFNLPCKSGMPELETIGLYKPYATTLG-DFM 68
Db 5 LIFLSFFGLCICFTALLRNQVYKNOKNLPFGTMGNPLREBTEIFLK-----LGSFM 58
Qy 69 QOHVSKYKTIYRSLNLFCEPTIVSADAGINREPLTONEGRLEFCSYPSRISGILGWSMLVL 128
Db 59 KNQARVSGFSPKSHLCPTIVSMDSELRNYLVNEAKGLVPGQSMIDILGKCNIAAV 118
Qy 129 VGDHNRDRLSISLFLNARLTLLKADVERHTLFVDSMOONSIFSAQDAKKTFTLM 188
Db 119 NSAKRYMKGALLSLISPTMRDQLLPKIDFPMKSHLTNW-DNKVIDIQETNNMAFLSS 177
Qy 189 AKHIMSMDPGEEET---EOLKKEYVTFMKGVASAPLPLPGTAHYKALQASRATILKFERK 245
Db 178 LKQI-----AGIESITSLAQEFMSEFFNLVGLSLPILNPNRYNRGQARKIYNLRTL 233
Qy 246 MEERKLDIKEDDQEEFEVKTEDEAEMSKSDHVRKQRTDDLLGVLSHNS--NLSTEQI 301
Db 234 IEERR-----ASKEIQHDMGLGYLMEAEATRFKLTDEM 266
Qy 302 IDLLSLFAGHETSSVALAIFLQACPAVEELREHELEIARAKKEGESEELNMDY 361
Db 267 IDLIITLLYSGYETVSTSMNAVYLLHDHPKVLDELRREHMAIREKKRP--EDPIDNDY 324
Qy 362 KKMFTQCVINETLRGLNVAVFLRKALKADVRYKGYDIPSGMKVLPVSAVHLNDSRYDQ 421
Db 325 RSMRFTRAVILETSRLATIVNGVLRKTTQDMENGIYIIPKMRIRIYVTRELNYPRLYPD 384
Qy 422 PNLFPNPMWQONNASSSGSGSFSTWGNMYMPFGGPRLCAGSELAKLEMAVEIHLVL 481
Db 385 PYSFNPMRMDKS-----LEHONSFLVFGGGRQCPGKELGVAEISTFLHAFVT 433
Qy 482 KFMNELAEDQPAFPEFVDPNGLPFRVSR 510
Db 434 KYRWEIIGDGLMKFPEVPEAPNGLRVRS 462

```

```

RESULT 3
C883_ARATH STANDARD: PRT: 490 AA.
ID C883_ARATH
AC 023051;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CYTOCHROME P450 86A3 (EC 1.14.-.-).
GN CYP86A3 OR YUP8H12.23.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Theologis A., Osborne B.I., Vysotskaia V.S., Federspiel N.A.,
RA Tortolmi M., Yu G., Oji O., Araujo R., Chung E., Dewar K., Dietrich F.,
RA Ecker J.R., Marzilli A., Oefner P., Davis R.W.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AC000098; AAB71462.1; -
CC InterPro: IPR001128; -
CC Pfam: PF00067; P450.1.
CC PRINTS: PR00385; P450.

```



DR PROSITE; PS00086; CYTOCHROME P450; 1.  
KW Oxidoreductase; Monooxygenase; Transmembrane; Heme.  
FT TRANSMEM 6 26 POTENTIAL.  
FT BINDING 439 439 HEME (BY SIMILARITY).  
SQ SEQUENCE 490 AA; 56409 MW; 7FD8CD7A8864D298 CRC64;

Query Match 21.6%; Score 579.5; DB 1; Length 490;  
Best Local Similarity 28.6%; Pred. No. 4.3e-30;  
Matches 133; Conservative 107; Mismatches 192; Indels 33; Gaps 9;

QY 34 KTRFNLPPGKSGWPLIGETIGYKPYTATTLGDFMQOHVSKYK--IYKSNLFGPTTVS 91  
DB 40 ENRHVLPDGLGMPITGNNMLSLRAFTSDPSFTRLIKRGPGYIKAHMFGNPSTIV 99  
QY 92 ADAGINRFLQNEGRLFECSYPRSTIGLIGKWSMLVVGDMHRMRSJSLNLSARLRT 151  
DB 100 TTSQTCRRVLTDD--AFKPGWPTSMELIGRKSFGVIGISEEHKRLRLTLAAVNGHEALS 158  
QY 152 ILKQVHHTLFLVDSMOONSIFSAODEAKKPTFNLMKHMSPGCEETQLKEYVT 211  
DB 159 TYIPTEENVITVLCKTKMGEEFLTLRLKLFRTI--MYIFLSESENVMDALREYTA 217  
QY 212 FMKGVASAPLNPCTAYHKALOSRAPIITLTERKMEERKLDIKEEDOEDEEVKTEDEAM 271  
DB 218 LNYGVRAVNPVGFAYHRAKAKRTIVAAPQSYTER--NQRKNLTKSKKDL 271  
QY 272 SKSDHYRQRDLDLGVNLKHSNLSTEQILDLLSLFAGHETSSVAIALAIFLQACP 331  
DB 272 DNLNVK-----DGD-----GKTLDEEIIIDVLLMYLNAGHESGHTIMATVFLQEH 320  
QY 332 KAVELREHLEIARAKKELESELNMDYKMDTQCVINETLRNGVNRRLHRAKLD 391  
DB 331 EYLQAKAKDEQELILKSRE--GQKGLSKETRMKELSOVYDDELTVTFSLAFRAKTD 379  
QY 392 VRYKGYDIPSGMKVLPVISAHLNDRYDQPNLFNPRMQOONNASSSGSGSFTWGN 451  
DB 380 VEMNGYLLPKGKVLTPWRDVIHDEVEFPDPKFPARM--DNG-----FVPKAGA 428  
QY 452 YMEFGGPRLCAGSELATLEMAVFIHHLVLFKFMWELAEEDQFAP 496  
DB 429 FLPEAGSHLCPGNDLAKLEISIFLHFLKLYQVRSNECEPVY 473

RESULT 4  
CP26\_BRAE STANDARD: PRT: 492 AA.  
AC P79739;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE CYTOCHROME P450 26 (EC 1.14.--) (RETINOIC ACID-METABOLIZING  
CYTOCHROME) (P450RA1) (RETINOIC ACID 4-HYDROXYLASE).  
GN CYP26.  
OS Brachydanio rerio (Zebrafish) (Zebra danio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Cyprinidae; Teleostei; Euteleostei; Ostariophysi;  
OC Cypriniformes; Cyprinidae; Rasbortinae; Danio.  
OX NCBI\_TaxID=7955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97094702; PubMed=8939936;  
RA White J.A., Guo Y.-D., Baetz K., Beckett-Jones B., Bonasoro J.,  
RA Hsu K.E., Dilworth F.J., Jones G., Petkovitch M.;  
RT "Identification of the retinoic acid-inducible all-trans-retinoic  
RT acid 4-hydroxylase.";  
RL J. Biol. Chem. 271:29922-29927(1996).  
CC -!- FUNCTION: PLAYS A KEY ROLE IN RETINOIC ACID METABOLISM. ACTS ON  
CC RETINIDS, INCLUDING ALL-TRANS-RETINOIC ACID (RA) AND ITS  
CC STEROISOMER 9-CIS-RA. CAPABLE OF 4-HYDROXYLATION, HYDROXYLATION,  
CC RESPONSIBLE FOR GENERATION OF SEVERAL HYDROXYLATED FORMS OF RA,  
CC INCLUDING 4-OH-RA AND 4-OXO-RA.  
CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.

CC -!- INDUCTION: BY RETINOIC ACIDS (RA).  
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL; U68234; AAC60045.1; -;  
DR ZFIN; ZDB-GENE-990415-44; cyp26.  
DR InterPro; IPR001128; -;  
DR Pfam; PF00067; P450.1;  
DR PRINTS; PR00385; P450.1;  
DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
KW Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;  
KW Endoplasmic reticulum.  
FT BINDING 438 438 HEME (BY SIMILARITY).  
SQ SEQUENCE 492 AA; 56281 MW; FD471435B2F30509 CRC64;

Query Match 18.8%; Score 503; DB 1; Length 492;  
Best Local Similarity 27.6%; Pred. No. 3.4e-25;  
Matches 144; Conservative 95; Mismatches 216; Indels 66; Gaps 14;

QY 7 HTLLPLLLPSLLSLFL-----ILKRRNRKTRNPPGSGWPLIGETIGYLP 58  
DB 4 YLLMTFLCQITLPLVPLFLAIAVLMELMIRRYDPCRSPLPGTGLPFGITGLD 61  
QY 59 YRATITLGMQOHVSKGYGIVNSLFGPTTVSADAGLNRLITONGRLFECSYPRSTIG 118  
DB 62 ---LQRRKFLRMKROQYGGCYVTHLEGNPTVRMGADNVOILLGKHLVSQWPAVST 118  
QY 119 ILGKSMVLVDMHRD-----MRSISLFLSHARLTLKDVHHTLFLVDSW--QONS 172  
DB 119 ILGSDTLNVHGVQKKNKKAIAMRAFSRALEH-----YIPVQGVASAIQEWLQKS 172  
QY 173 TFSADAKKFTFNLMKHMSPGCEET--EQLKEYVTFMKGVASAPLNPCTAYH 230  
DB 173 CVLVPEMKKLMFRIMARLITLGFEPQITDQELVAFEEEMKKNLPIIDVPEGLY 232  
QY 231 ALQSRATILKTERKMEERKLDIKEEDOEDEEVKTEDEAMSDHYRQRDLDLGV 290  
DB 233 GLRAR---NFIHSKLEENIRKKIODDNEQKXKDALQL---IENSRSD----- 279  
QY 291 LKHSNLSTEQILDLLSLFAGHETSSVAIALAIFLQACPRAVEELREHLEIARAKE 350  
DB 280 ----PESLOAKMEATELLFGGHETTASTATSLVMTGLNTEVYQVRE---VQKYE 331  
QY 351 LG---ESELNMDYKMDTQCVINETLRNGVNRFLHRAKLDVRYKYDIPSGMKV 406  
DB 332 MCMYTPGKGLSMELLDQLKYTCVIKETLRINPVPVGGFRVALKTFELNGYQIPKGMNY 391  
QY 407 PYSANHLNDRYDQPNLFNPRMQOONNASSSGSGSSTNGNNMPPGGRPLCAGSE 466  
DB 392 YSICDTHADVAPVFNKEEPQPERFM-----SKIGEDSGR--NIPPGGSRMCVKE 442  
QY 467 LAKLEMAVFIHHLVLFKFMWELAEEDQ-----PFAFPVVP 502  
DB 443 FAKVLKILTVLVLQHCNNILNSGPPMTKGTPIYVVDMLP 483

RESULT 5  
C881\_MAIZE  
ID C881\_MAIZE STANDARD: PRT: 519 AA.  
AC 043246;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE CYTOCHROME P450 88A1 (EC 1.14.--) (DMARF3 PROTEIN).  
GN CYP88A1 OR D3.

```

OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae;
OC Andropogoneae; Zea.
NCBI_TaxID=4577;
RN [1]
RC STRAIN=CV. B73;
RX MEDLINE=96004534; PubMed=7549486;
RA Minkler R.G., Helentjaris T.;
RT "The maize Dwar3 gene encodes a cytochrome P450-mediated early step
RT in gibberellin biosynthesis."
RL Plant Cell 7:1307-1317(1995).
CC -1- PATHWAY: EARLY STEP IN GIBBERELLIN BIOSYNTHESIS.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ROOTS, DEVELOPING LEAVES, THE
CC VEGETATIVE MERISTEM, AND SUSPENSION CULTURE CELLS.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32579; AAC49067.1; -
DR InterPro: IPR001128; -
DR Pfam: PF00067; P450.1;
DR PROSITE: PS00086; CYTOCHROME_P450.1;
KW Oxidoreductase; Monooxygenase; Transmembrane; Heme.
FT BINDING 1 21 POTENTIAL.
FT SEQUENCE 519 AA; 57906 MW; 0F8977A024316D95 CRC64;

```

Query Match 18.3%; Score 491.5; DB 1; Length 519;  
 Best Local Similarity 28.2%; Pred. No. 2e-24;  
 Matches 129; Conservative 91; Mismatches 189; Indels 49; Gaps 11;

```

QY 36 RFLNPGSGSPFGETIGYIKFTATTLSDPMOONHYSKGR--IYRNLGEPPI-USA 92
DB 69 RARLPGGMGRLVGGWMAFLRAKSGKPDFAISFVRFRTGVRSGFMSSPVLVTT 128
QY 93 DAGLNRFILNEGRLEFECYSRISGILGKMSLVLVDMRDKMSISLNFSLHARLRTI 152
DB 129 AEGCKQVLMDDA---FTYGMKATVALVGRPSYAMPDERIRKRLKTAALINGFDALTG 186
QY 153 LKQVERHTLFLVDSW--QONSIFSAODEAKKFTNLAKHIMSDPGEETEOLKEEYVT 211
DB 187 YLPFDRTVYSSLRAMDHGSGVEFLTELRRMTFKIIYO--IFLGADQATTRALERSYTE 245
QY 212 FMKGVNSAPLRLPGTAHKAALOSRATILKFERKMEERKLDIKE-----EDQEEEEKYT 265
DB 246 LNYGRRRAALVLPFGATRGALRRRLVAALQVLDERRAARAGVSGGVDMMDRLIEA 305
QY 266 EDEAEMSKSDHVRQRDDDLGLWVLKHSNLSTEOILDLISLFAGHETSVAIALAIF 325
DB 306 QDE-----RGRHLDD-----EIIDVLVMYNLAGHSSGCHITMATV 342
QY 326 FLQACPRAVELREHELEIARAKKEGESELMNDYKMDFTQCVINETLGLVNVRLH 385
DB 343 FLQENPDMFARAKAEQEAIMRSIPS--SQRGTLRLDFRKMELISOVIDETLRLVINSFSF 401
QY 386 RKALKDVRYKGYDIPSGKVLPIVSAVHLDSRYDQPLTFPMWMOONNASSSGSSSF 445
DB 402 RQARFDVFNQYLPLPKGKQVLMWRSVHMDQVYDPDKFKPSRWE-----GHSBRAGTF 456
QY 446 STWGNMYPEGGCPRLCAGSELAKLEMAVFIHHLVLF 483
DB 457 -----LAFGLGARLCGNDLAKLEISVFLHFLLCG 487

```

```

RESULT 6
CP26_MOUSE
ID CP26_MOUSE STANDARD: PRT: 497 AA.
AC 055127;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CYTOCHROME P450 26 (EC 1.14.-.-) (RETINOIC ACID-METABOLIZING
DE CYTOCHROME) (P450RA1) (RETINOIC ACID 4-HYDROXYLASE).
CN CYP26A1 OR CYP26 OR P450RA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
NCBI_TaxID=10090;
RN [1]
RC STRAIN=CV.
RX MEDLINE=97392446; PubMed=9250660;
RA Fujii H., Sato T., Kaneko S., Gotoh O., Fujii-Kuriyama Y., Osawa K.,
RA Kato S., Hamada H.;
RT "Metabolic inactivation of retinoic acid by a novel P450
RT differentially expressed in developing mouse embryos."
RL EMBO J. 16:4163-4173(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98113212; PubMed=9442090;
RA Abu-Abed S.S., Beckett B.R., Chiba H., Chithalen J.V., Jones G.,
RA Metzger D., Chambon P., Petkovich M.;
RT "Mouse P450RA1 (CYP26) expression and retinoic acid-inducible retinoic
RT acid metabolism in F9 cells are regulated by retinoic acid receptor
RT gamma and retinoid X receptor alpha."
RL J. Biol. Chem. 273:2409-2415(1998).
CC -1- FUNCTION: PLAYS A KEY ROLE IN RETINOIC ACID METABOLISM. ACTS ON
CC STEROISOMER 9-CIS-RA. CAPABLE OF BOTH 4-HYDROXYLATION AND 18-
CC HYDROXYLATION. RESPONSIBLE FOR GENERATION OF SEVERAL HYDROXYLATED
CC FORMS OF RA, INCLUDING 4-OH-RA, 4-OXO-RA, AND 18-OH-RA.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.
CC -1- INDUCTION: BY RETINOIC ACIDS (RA).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y12657; CA73206.1; -
DR MGD: MGI:1096359; CYP26.
DR InterPro: IPR001128; -
DR Pfam: PF00067; P450.1;
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450.1;
KW Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
KW Endoplasmic reticulum.
FT BINDING 442 442 HEME (POTENTIAL).
FT SEQUENCE 497 AA; 56177 MW; 33B07D7C29134471 CRC64;

```

Query Match 17.2%; Score 461; DB 1; Length 497;  
 Best Local Similarity 28.4%; Pred. No. 1.6e-22;  
 Matches 148; Conservative 78; Mismatches 231; Indels 64; Gaps 14;

```

QY 10 LPLLLPSL-----LSLFLFLK-----RRNRKTRNLPPGKSGWPLGFTIGYLK 57
DB 3 LPALLASALCTFVLPPLFLFLAALKIMDLVYSSRDSGALPLPGTWGMPFGETL---- 58
QY 58 PYRTTTLGDEMQHVSXGKTYRNLGEPPIVSADAGLNFIITQNBRLFECSYPSRISG 117
DB 59 -QWVLRKRRFLQMKRRRYGFTYKTHLFGCPVRYWAGADVNRRIILGELHLYSVHMPASVR 117

```

QY 118 GILCKMSLVGVGMHMRMSISINFLSHARLRTLLKDVNHTLVLDWMQ--NSIFS 175  
 DB 118 TILGAGCLSNLHDSHKKRKVIMQAFSREALQCVLV-IAEVSQCLEOMLSCGERGL 176  
 QY 176 AODEAKKFTFLMAKHIMSMPGE-----EETEOLKEEYVTMKGVSAPLNPGIAYKA 231  
 DB 177 VYPEKRLMFRIARILLKILGCEPAGGEDOOLVEAFEMTRNLFSLPIDVPFSGLYRG 236  
 QY 232 LOSRATILKFERKME--RKLDIKEED--OEVEEVKTEDEAEMSKSDHVKOITDDL 286  
 DB 237 VKARLILHARIEENIRAKIRLQATERPGCKDALQILIEHSWE-----RGERLDMQ- 288  
 QY 287 LGWVLKHSNLSTEOILDLILSLFAGHETSSVAIALAIFLOACPRAVEELREHLEIAR 346  
 DB 289 --ALKOS--STE-----LLFGHETSAATSLLTYIGLVPHVLOKVRKEELKSKGL 335  
 QY 347 AKKEGESELMWDYKKMDFQCVINETLRIGNVRFILHRKALKVRYKGDIPSGMKVL 406  
 DB 336 LCKSNODKLMETLEOLKTYIGCVIKETRLNPPVPGGFRAALKTFELNGVQIPKGMVVI 395  
 QY 407 PVISAVHLNDSRYDQPNLFNPMWMOQNNGASSSGSFSFWNGNMPFGGPRICAGSE 466  
 DB 396 YSICDTHVADIFINKEEFNDRFIVPHEDASRPS-----FIFPGGLRSCVKE 446  
 QY 467 LAKLEMAVFIHHLVLFKFWMLAE-----DDQPFAPFPDFP 502  
 DB 447 FAKILKIFTEVLARHCDWOLLNGPPTKTSPTYVPVNDLP 487  
 RESULT 7  
 CP26\_HUMAN STANDARD; PRT: 497 AA.  
 AC 043174;  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 40, Last annotation update)  
 DE CYTOCHROME P450 26 (EC 1.14.--) (RETINOIC ACID-METABOLIZING  
 DE CYTOCHROME) (P450RA1) (RETINOIC ACID 4-HYDROXYLASE).  
 GN CYP26A1 OR CYP26.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97373542; PubMed=9228017;  
 RA White J.A., Beckett-Jones B., Guo Y.-D., Dilworth F.J., Bonasoro J.,  
 RA Jones G., Petkovich M.;  
 RT "cDNA cloning of human retinoic acid-metabolizing enzyme (hP450RA1)  
 RT identifies a novel family of cytochromes P450.";  
 RL J. Biol. Chem. 272:18538-18541(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98380037; PubMed=9716180;  
 RA Sonneveld E., van den Brink C.E., van der Leede B.M., Schultes R.K.,  
 RA Pekovich M., van der Burg B., van der Saag P.T.;  
 RT "Human retinoic acid (RA) 4-hydroxylase (CYP26) is highly specific for  
 RT all-trans-RA and can be induced through RA receptors in human breast  
 RT and colon carcinoma cells.";  
 RL Cell Growth Differ. 9:629-637(1998).  
 RN [3]  
 RP TISSUE SPECIFICITY.  
 RX MEDLINE=99045433; PubMed=9826557;  
 RA Trofimova-Griffin M.E., Juchan M.R.;  
 RT "Expression of cytochrome P450RA1 (CYP26) in human fetal hepatic and  
 RT cephalic tissues.";  
 RL Biochem. Biophys. Res. Commun. 252:487-491(1998).  
 CC -1- FUNCTION: PLAYS A KEY ROLE IN RETINOIC ACID METABOLISM. ACTS ON  
 CC RETINOIDS, INCLUDING ALL-TRANS-RETINOIC ACID (RA) AND ITS  
 CC STEROISOMER 9-CIS-RA. CAPABLE OF BOTH 4-HYDROXYLATION AND 18-  
 CC HYDROXYLATION. RESPONSIBLE FOR GENERATION OF SEVERAL HYDROXYLATED  
 CC FORMS OF RA, INCLUDING 4-OH-RA, 4-OXO-RA, AND 18-OH-RA.  
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.

CC -1- TISSUE SPECIFICITY: HIGHEST LEVELS IN ADULT LIVER, HEART,  
 CC PITUITARY GLAND, ADRENAL GLAND, PLACENTA AND REGIONS OF THE BRAIN.  
 CC -1- INDUCTION: BY RETINOIC ACIDS (RA).  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF005418; AAB88881.1; -  
 DR MIM: 602239; -  
 DR InterPro: IPR001128; -  
 DR Pfam: PF00067; P450; 1.  
 DR PRINTS: PR00385; P450.  
 DR PROSITE: PS00086; CYTOCHROME\_P450; 1.  
 KW Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;  
 KW Endoplasmic reticulum.  
 FT BINDING 442 442 HEME (POTENTIAL).  
 SQ SEQUENCE 497 AA; 56162 MW; EAB6B84B24B2EAB3 CRC64;  
 Query Match 16.3%; Score 438; DB 1; Length 497;  
 Best Local Similarity 27.0%; Pred. No. 4.9e-21;  
 Matches 142; Conservative 85; Mismatches 226; Indels 72; Gaps 15;  
 QY 10 LPALLPSL-----LSLFLILK-----RRNRKTRFNLPKSGMPLGETIGYLK 57  
 DB 3 LPALLASALCTVPLPLFLFLAIKLMDLYCVSGRSCSLPLPGMSPFFGETL----- 58  
 QY 58 PYTATLGDPMQOAHVSKIGKIRSNLFGPEPTVSADAGLNRILONEGFLFCSPRSTG 117  
 DB 59 -OMVLORRFFLOMKRRKGYFTKTHLFGPRVVGADVNRILLGDDRLVSVMHPASVR 117  
 QY 118 GILCKMSLVGVGMHMRMSISINFLSHARL---TILKVERHTLVLDWMQ--NS 172  
 DB 118 TILGAGCLSNLHDSHKKRKVIMQAFSREALQCVLV-----LQWLSGGER 173  
 QY 173 IPSAODEAKKFTFLMAKHIMSMP-----GEETEOLKEEYVTMKGVSAPLNPGTA 227  
 DB 174 GLVVPYVKRLMFRIARILLKILGCEPQLAGDQSE-QQLVEAFEMTRNLFSLPIDVPFSG 232  
 QY 228 YKALQSRATILKFERKMEERKLDIKEED-----OEVEEVKTEDEAEMSKSDHVKOIT 282  
 DB 233 IYRGMKARLILHARIEONIRAKICGLRASEAGCGCKDALQILIEHSWE-----RGERL 285  
 QY 283 DDDDLGWLKHSNLSTEOILDLILSLFAGHETSSVAIALAIFLOACPRAVEELREHIL 342  
 DB 286 DMQ-----ALKOS--STE-----LLFGHETSAATSLLTYIGLVPHVLOKVRKEELK 331  
 QY 343 ETARAKKEGESELMWDYKKMDFQCVINETLRIGNVRFILHRKALKVRYKGDIPSG 402  
 DB 332 SKGLLCKSNODKLMETLEOLKTYIGCVIKETRLNPPVPGGFRAALKTFELNGVQIPKG 391  
 QY 403 WKVLPIVSAVHLNDSRYDQPNLFNPMWMOQNNGASSSGSFSFWNGNMPFGGPRIC 462  
 DB 392 WNVVISICDTHVADIFINKEEFNDRFIVPHEDASRPS-----FIFPGGLRSC 442  
 QY 463 AGSELAKLEMAVFIHHLVLFKFWMLAE-----DDQPFAPFPDFP 502  
 DB 443 VKKEPAKILKIFTEVLARHCDWOLLNGPPTKTSPTYVPVNDLP 487  
 RESULT 8  
 YRV2\_CAEEL STANDARD; PRT: 520 AA.  
 ID YRV2\_CAEEL  
 AC Q27514;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)



```

OY 289 WLK-HSMLSTFQIIDLILSLFAGHETSSVAIALAIFLOACPRAVELEHREHLEIARA 347
D 303 EVLVKDKHLEIFIIQIOLFEIILAGYDTTALISYSSYLATHEPQIOKLOE-----V 356
OY 348 KKEIGESLMDDDKKMFOCVINETRLGNVAFELH-RAKAKYRKRGKDIPSGMKVL 406
D 357 DRECPDPFVTDQSLKLYLEECVAKELRLPLASLVHNRCLKTNTVIGMEIEGTIN 416
OY 407 PVISAVHLDNRY-DQPLFNPMMQOONNGASSSGSSEFTWGNMYPGGPRCAGS 465
D 417 VDTMSLHNDPRKMGDVAEFPKRE-----SGDELFFAKG-GYLPFGKPRICIGM 467
OY 466 ELARLEMAVFIHHLVLEKNWE 486
D 468 RLAMMEMKMLTNILKNYTFE 488

RESULT 10
C912-ARATH STANDARD; PRT: 500 AA.
AC 065790: 09S2U3:
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DE CYTOCHROME P450 91A2 (EC 1.14.-.-).
GN CYP91A2 OR AT4G37430 OR PG17.80.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=CV_COLUMBIA; TISSUE=Seedling;
RX MEDLINE=98281573; PubMed=9620263;
RA Mizutani M., Ward E., Ohta D.:
RT "Cytochrome P450 superfamily in Arabidopsis thaliana: isolation of
RT cDNAs, differential expression, and RFLP mapping of multiple
RT cytochromes P450."
RL Plant Mol. Biol. 37:39-52(1998).
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN=CV_COLUMBIA;
RX MEDLINE=20083488; PubMed=10617198;
RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA Poll T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA Harris B., Anstorge W., Brandt P., Grivell L., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
RA Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidheini T.,
RA Reichert B., Portellelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA Vos P., Hohenseil J., Zimmermann W., Medler H., Ridley P.,
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA Van der Schueren J., Grymonprez B., Chung Y.-J., Vandenbussche F.,
RA Braeken M., Welfjens I., Voet M., Bastlaens I., Aert R., Defoor E.,
RA Holtzenger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA Holzer E., Brandt A., Peters S., van Staveren M., Dikse W.,
RA Moolman P., Klein lankhorst R., Rose M., Hauf J., Koeter P.,
RA Benelaiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA De Keijser A., Blyssnaet C., Gielen J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., Molay K., Mayes R.,
RA Petlett A., Rajandream M.-A., Lyne M., Benes V., Rechmann S.,
RA Bokkova D., Blocker H., Scharfe M., Grimm M., Loehner T.-H.,
RA Dose S., de Haan M., Maarse A., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fatmann B., Granderath K., Dauner D., Herzl A.,
RA Neumann S., Argitlou A., Vitale D., Liguori R., Piravandi E.,
RA Massenot O., Quigley F., Glabaud G., Muendlein A., Felber R.,
RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA Chefor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA Gibbons T., Weber N., Vandenbol M., Bagues M., Terol J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,
RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,

```

```

RA Zaccaria P., Beyan M., Wilson R.R., de la Bastide M., Habermann K.,
RA Pannell L., Dedha N., Gnoj L., Schütz K., Huang E., Spiegel L.,
RA Sakem M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Mix P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kimer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA Nelson J., Spith J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,
RA Granat S., Shohdy N., Hasegawa A., Hamed A., Lodhi M., Johnson A.,
RA Chen E., Marra M., Martienssen R., McCombie W.R.:
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT thaliana."
RL Nature 402:769-777(1999).
CC -i- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: D78607; BAA28539.1; -
DR EMBL: AL035601; CAB83210.1; -
DR EMBL: AL161591; CAB80408.1; -
DR Mendel: 29894; Arath:1113.29894.
DR InterPro: IPR001128; -
DR Pfam: PF00067; P450.1.
DR PROSITE: PS00086; CYTOCHROME_P450.1.
KW Oxidoreductase; Monooxygenase; Membrane; Heme.
FT BINDING 438 438 HEME (BY SIMILARITY).
FT CONFLICT 106 106 A -> V (IN REF. 1).
FT CONFLICT 127 127 I -> M (IN REF. 1).
FT CONFLICT 140 140 N -> I (IN REF. 1).
FT CONFLICT 454 454 S -> T (IN REF. 1).
SQ SEQUENCE 500 AA; 57555 MW; 0FB453D2070EA2EA CRC64;

Query Match 11.8%; Score 316; DB 1; Length 500;
Best Local Similarity 25.2%; Pred. No. 3.1e-13;
Matches 135; Conservative 83; Mismatches 199; Indels 118; Gaps 22;

OY 9 LPLLLPSLILFLILKLR--RNKRTFNLPPGKSGMPLGFTIGLYKPYATTLGD 66
D 1 MLYFLLP-----LFLIVSYKFLSKQRFNLPPGSPRPVGH-LHLMKP-----PIHR 50
OY 67 FMOQHVSKYKTKYSNLFGEPTIV-----SADAGINRFILONEGRLEFCSYPRSTG 117
D 51 LLORYSNDYGPFL-SLRGSRVRVITSPSLAOESFTGNDIVLSRPLQTLAKVAVYAH 109
OY 118 GLTGKMSLVLVGDHMRMRST-SLNPFSHARLRT--LLKVENHTLFEVDSMOONSIF 174
D 110 TTVG-----TAPYGDWRNLRICSOELISHRILNFQHRKQELIRMLRSLRYQTS-- 163
OY 175 SAODEAKRTF-----NLMAKHMSMDP-----EETEDOLKKEVYFMKV 216
D 164 ---NSNDFTHLELRPLSLDFNNIVRVYKRYGDDVNNKEALELKK----- 211
OY 217 VSAPLNFGTAVHKLQSRAT-----LKTIEKMERKLDIEEDOEVEVYTEDEA 269
D 212 -----LVYDIDAMYSGANHSADYPLKLFKGNKFEKVAIG-----KSMDDI 253
OY 270 ESKSDHVKORTDDDLGWLKHSNLSTFQILD-----LILSLFAGHETSSVAIALAI 324
D 254 LQRLDECRDRKESGTAVNHLISLQDOQPEYTYVITIGLMSHMLAGTETSAYVLEMAN 313
OY 325 FFLQACPRAVELEHLEIARAKKEIGESSEL--NMDYKMDFTQCVINETRLGNVAF 383

```

Db 314 ANLRNPEVLEKARSE-----IDKIGKRLIDESDIAVLPLQNVSETRFLPVPAPF 367  
 OY 384 L-HRRALKDVRYKGYDIPSGMKVLPVISAVALHDSNRYPDPLFNPWMOQONGASSSGS 442  
 Db 368 LIPSPIDDMKIGCYDVRDRTIVWAMAIHRDEIWEPEKFPNDRK-----NDGC----- 419  
 OY 443 GSFSTGWNVY-----MPFGGPRILCAGSELAKLEMAVFIHHLVLKFMWELAEDDO 492  
 Db 420 -----GSDYVYVKMPFGNGRRTCPGALGORIVTALGSLICFEWENKGE 468

RESULT 11  
 CP32\_RAT STANDARD: PRT: 504 AA.  
 ID CP32\_RAT 064672: 064629:  
 AC P05183: 064672: 064629:  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE CYTOCHROME P450 3A2 (EC 1.14.14.1) (CYP11A2) (P450-PCN2) (P450/6-BETA-7) (TESTOSTERONE 6-BETA-HYDROXYLASE).  
 GN CYP3A2 OR CYP3A-2.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Rattus. NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-87064606: PubMed=3785219:  
 RA Gonzalez F.J., Song B.-J., Hardwick J.P.:  
 RT "Pregnenolone 16 alpha-carbonitrile-inducible P-450 gene family: gene conversion and differential regulation.";  
 RL Mol. Cell. Biol. 6:2969-2976(1986).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SPRAGUE-DAWLEY: TISSUE=Liver:  
 RX MEDLINE-91254339: PubMed=2043144:  
 RA Miyata M., Nagata K., Yamazoe Y., Kato R.:  
 RT "A gene structure of testosterone 6 beta-hydroxylase (P450111A).";  
 RL Biochem. Biophys. Res. Commun. 177:68-73(1991).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SPRAGUE-DAWLEY: TISSUE=Liver:  
 RX MEDLINE-95070139: PubMed=7979376:  
 RA Miyata M., Nagata K., Shimada M., Yamazoe Y., Kato R.:  
 RT "Structure of a gene and cDNA of a major constitutive form of testosterone 6 beta-hydroxylase (P450/6 beta A) encoding CYP3A2: RT comparison of the cDNA with P450PCN2.";  
 RL Arch. Biochem. Biophys. 314:351-359(1994).  
 CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.  
 CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) -> ROH + OXIDIZED FLAVOPROTEIN + H(2)O.  
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.  
 CC -1- INDUCTION: BY PREGNENOLONE 16-ALPHA-CARBONITRILE (PCN2).  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M13646: AAA41051.1: -  
 DR EMBL: X79319: CAAS5887.1: -  
 DR EMBL: X79320: CAAS5888.1: -  
 DR EMBL: U09742: AAA82168.1: -  
 DR EMBL: U09734: AAB60492.1: -  
 DR EMBL: U09725: AAB60492.1: JOINED.

DR EMBL: U09726: AAB60492.1: JOINED.  
 DR EMBL: U09727: AAB60492.1: JOINED.  
 DR EMBL: U09728: AAB60492.1: JOINED.  
 DR EMBL: U09729: AAB60492.1: JOINED.  
 DR EMBL: U09730: AAB60492.1: JOINED.  
 DR EMBL: U09731: AAB60492.1: JOINED.  
 DR EMBL: U09732: AAB60492.1: JOINED.  
 DR EMBL: U09733: AAB60492.1: JOINED.  
 DR PIR: A25222: A25222.  
 DR HSSP: P14779: IFAG.  
 DR InterPro: IPR001128: -  
 DR InterPro: IPR002397: -  
 DR InterPro: IPR002401: -  
 DR InterPro: IPR002402: -  
 DR InterPro: IPR002403: -  
 DR Pfam: PF00067: P450.1.  
 DR PRINTS: PR00359: BP450.  
 DR PRINTS: PR00385: P450.  
 DR PRINTS: PR00463: EP450I.  
 DR PRINTS: PR00464: EP450II.  
 DR PRINTS: PR00465: EP450IV.  
 DR PROSITE: PS00086: CYTOCHROME\_P450: 1.  
 DR Oxidoreductase, Monooxygenase, Electron transport, Membrane, Heme; KW Microsome, Endoplasmic reticulum.  
 FT BINDING 443 443 HEME.  
 FT CONFLICT 56 57 GL -> AV (IN CAAS5887).  
 FT CONFLICT 429 429 D -> H (IN REF. 1).  
 FT CONFLICT 445 445 G -> D (IN REF. 1).  
 SQ SEQUENCE 504 AA: 57731 MW: C2C84AF736035AB2 CRC64;

Query Match 11.7%; Score 312.5; DB 1; Length 504;  
 Best Local Similarity 25.5%; Pred. No. 5.2e-13;  
 Matches 137; Conservative 96; Mismatches 211; Indels 93; Gaps 24;

OY 9 LPLLLPSLSLFLILKR--RNKTRNLP--PEKSGMPLGELIGYLYKPYTATL 64  
 Db 3 LLSALTLETWLVLLVILVLRGLGTHRGIGKKGIDPCKPLPLGTVLNYK-----GL 57  
 OY 65 GDFMOQHVSKYKGIYRSNLFGPEPTIVSA--DAGLNRFILQNEGRLEFC---SYPRSIG- 117  
 Db 58 GRFMECKKTKYKTIW--GLFDGQTPVPAIMDEMIKNVL-----VKECFSVFTNRDGP 110  
 OY 118 -GILKSMVLVGDHMDMSISLNLNARLRTILKLDVRRHTLFLVDSMOONS---- 172  
 Db 111 VGINGKAASVAKDEMKRYRALLSPFTFS-GRLKE-MPIIEQGDILVKYLKQEAETGK 168  
 OY 173 -----IFSA--QDEAKKFTFNMAKHLMS--MDPGEETEDLKR--EYVTFMKGVSA 221  
 Db 169 PVTMKVFGAYSMDVITSTSGVAVDSLNNPKDPFEVETKKLRFDFEDPLFLSVLFPF 228  
 OY 222 NLPGTAHVHKL-----QSRATILKFERKMEERKLDIKEEDDEBEVKTDEAEKMS 274  
 Db 229 LTP---IYEMLNTGMFPKDSIAFQKFEVHR-ITKRLDSKKHR-----VDPLQMLN 277  
 OY 275 DHRVKORTDDLLGVMLKHSNLSTEQILDLILSLFAGHETSVAIALAIFLOACPRAV 334  
 Db 278 AH-----NNSKDE-----VSHKRLSDVEIIAQSIVIFIFAGYETSTSLSVLFLATH 329  
 OY 335 EELREHELEIARAKKEIGSELNMDYKMDQTCVNETRLGNAVVFRLRKALKDVR 394  
 Db 330 KRLQEE-IDGALPSK-----APPTVDIWMEMEYLDVNETRLXLPDGLRLEKVCADTEL 384  
 OY 395 KGYIIPSGMKVLPVISAVALHDSNRYPDPLFNPWMOQONGASSSGSSTGNNYMP 454  
 Db 385 DGLIIPGSGVVTIPTVALHNDPQHPMPKEEFHPRFSEKNNKSIDP-----YYVLP 435  
 OY 455 FGCGPRILCAGSELAKLEMAVFIHHLVLKFMWELAEDDOOPFAFPVDPNGLPINRSR 511  
 Db 436 FGNGPRNCIGMRFLAMNKKLITVLOVNFSPQCKETQ-----IPLKLSR 480

RESULT 12



```

Db 140 SLRKIMSTVEETVEELMHLDASAKRAVDLLDYQE-----FTLIDIGRIA 187
Oy 193 MSMDEGEET-----EOLKKEVTFPMKG-----VYSAPLN.PGT-----AYHKAL 232
Db 188 M-----GQIESLMEFRNPMPLPKKEIFKKGKMPFLIAGVFPILGILMRLMKFPKFSFAF 243
Oy 233 OSRATILFKFERKMEERKLDIKE-----EDQEEVEKTEDEAEMSKSDHYRKORTDDL 286
Db 244 GIMTMEKALNRLEQRAADKAGIEPSEPODFIDFLDARANY---DFIEEST----- 296
Oy 287 LGM-----VLK-HSNLSTQIIDLILSLFAGHETSSVAIALAIFLQACPRAVEELREEH 341
Db 297 LGFAKSEVLAKVDKHLTFPEIIGQLFVFLIAGVDTLALSLSSYLLATHNPETIOKRLQEE- 355
Oy 342 LEIARAKKELGESLMDYKMKMFTOCVINEITRLGNVRFILH-RKALKDVRKGYDIP 400
Db 356 -----VDECEPDPEVTFQIOISKLMCEVKEALMRYPLASLHNRKCKMTNVLGVEID 410
Oy 401 SGWKVLPVISAVALHLDNSRY-DQPNLFNPMRQOONNGASSSGSFSFGWNNYMPFGGCP 459
Db 411 EGTNVQVDTWTLHVDPKYMGDASSEKPERME-----TGDELFLYAKG-GYLFPGMGP 461
Oy 460 RLCAGSELAKLEMAVFIHHLVLEKEMLEDDQ 492
Db 462 RICIGMRLAMMEKLLTHILKKTFTSTETE 494

```

```

RESULT 14
CP36_MOUSE STANDARD: PRT: 501 AA.
ID CP36_MOUSE
AC 054750;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CYTOCHROME P450 2J6 (EC 1.14.14.1) (CYPLI6) (ARACHIDONIC ACID
  EPOXYGENASE).
GN CYP2J6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Liver;
RX MEDLINE=98234557; Pubmed=9570962;
RA Ma J., Ramachandran S., Fiedorek F.T. Jr., Zeldin D.C.;
  "Mapping of the CYP2J cytochrome P450 genes to human chromosome 1 and
  mouse chromosome 4."
RT Genomics 49:152-155(1998).
RL -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
  OXIDIZED FLAVOPROTEIN + H(2)O.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U62295; AAB87636.1; -
DR MGD: MGI:1270148; Cyp2j6.
DR InterPro: IPR001128; -
DR InterPro: IPR002401; -
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PRINTS: PR00463; EP4501.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
DR OXidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
  KM Microsome; Endoplasmic reticulum.

```

```

FT BINDING 447 447 HEME (BY SIMILARITY).
SQ SEQUENCE 501 AA: 57820 MM; 4E5A26B35D831E37 CRC64;
Query Match 11.3%; Score 303; DB 1; Length 501;
Best Local Similarity 23.2%; Pred. No. 2,le-12;
Matches 126; Conservative 98; Mismatches 204; Indels 116; Gaps 21;
Oy 9 LPLLLPSLLSLFLFLILKRRNKRTPFNLPKGSMPPIGETIGYLKPTATATLGGFM 68
Db 16 LHPRLVLAATVFLFLALDYFKNRKPK---NYPGGMGLPFGNIFOL-----DFG 62
Oy 69 QOHS-----KYGRTYRNL-----FGEPTVSADAGLNREPLONEGRLECSYPR 114
Db 63 QPHLSIOPLVKYGNIFFSLNGDITSVITGGLPLKEALTOGQONIMR-----PL 113
Oy 115 SI--GGILGKSMVLVVDMDHMRSLNLSARL-RTILKDVENHTLFLVDSMOON 171
Db 114 SVMQERISNKNGLIFSSQIWKQORFALMTLRLNGLCKSLSEERMQDEASHLVEAIR-- 171
Oy 172 SIFSADAKKFT-----FNLMKIHMSMDGE-----EETOLKKEVTFPMKGIV 217
Db 172 -----EEGKFPNPFSTNNVSNITCSVTGERFDYDSRFQEMRLDDEVYLETMTI 226
Oy 218 SAPLN-----LPGTAYHRALOSRAITLKFIERKMEERKLDIKEEDQEEVEKTEDEA 269
Db 227 SOLYNIFFPMIKKIYIPG-SHOKVFRWMEKLFVSCMIDHR-----KQMPNDEPRDFIDA 280
Oy 270 ---EKSQSDHVRKORTDDLLGWLKHSNLSLEQIIDLILSLFAGHETSSVAIALAIF 326
Db 281 FLKEWTK-----YPEKTSFNEENLICSTLDFAGTETSTTLFMALLY 325
Oy 327 LQACRAVEELREHLEIARAKKELGESLMDMDYKMKMFTOCVINEITRLGNVRF-LH 385
Db 326 MALPVEQEKYQAEIDRYIGQR-----AARLADRESPTNNAVYHEKRGNTIPLNVP 380
Oy 386 RKALKDVRKGYDIPSGMKVLPVISAVALHLDNSRYDQPNLFNPMRQOONNGASSSGSFS 445
Db 381 REVAMDTNLNFPILPKGMVLTNLALHDPKEMATPVFNPHELE-----NQGF 431
Oy 446 STWGNMYMPFGGCPRLCAGSELAKLEMAVFIHHLVLEKEMLEDDQPAFPVDPFN 503
Db 432 KK-RESFLPFSMGKRAKGLGEOLARSELFIPTSLMKFTFPPIEKLSP-----KFRN 484
Oy 504 GLPI 507
Db 485 GLTL 488

```

```

RESULT 15
CPA8_MESAU STANDARD: PRT: 494 AA.
ID CPA8_MESAU
AC P24434;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CYTOCHROME P450 2A8 (EC 1.14.14.1) (CYPLI8) (P450-MC1) (P450-AFB).
GN CYP2A8.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
  Mesocricetus.
NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=91112759; Pubmed=2275554;
RA Lai T.S., Chiang J.Y.L.;
  "Cloning and characterization of two major 3-methylcholanthrene
  inducible hamster liver cytochrome P450s."
RT Arch. Biochem. Biophys. 283:429-439(1990).
RL [2]
RN SEQUENCE FROM N.A.

```



```

RC TISSUE=Liver;
RX Eukubara M., Nagata K., Mizokami K., Yamazoe Y., Takanaka A., Kato R.;
RA "Complete cDNA sequence of a major 3-methylcholanthrene-inducible
RT cytochrome P-450 isozyme (P-450AEB) of Syrian hamsters with high
RT activity toward aflatoxin B1."
RL Biochem. Biophys. Res. Commun. 162:265-272(1989).
CC -1- FUNCTION: HIGHLY ACTIVE IN 7-ETHOXYCOUMARIN O-DEETHYLATION, AND
CC BENZHEPHTHENE N-DEETHYLATION; MODERATELY ACTIVE IN TESTOSTERONE
CC 7-ALPHA-HYDROXYLATION, ETHYLORPHINE N-DEETHYLATION, P-
CC NITROANISOLE O-DEETHYLATION, AND ONLY SLIGHTLY ACTIVE IN
CC BENZOPHYRENE 3-HYDROXYLATION, 7-ETHOXYRESORUFIN O-DEETHYLATION,
CC TESTOSTERONE 2-ALPHA-HYDROXYLATION AND TESTOSTERONE 17-OXIDATION.
CC COMPETENT IN THE METABOLIC ACTIVATION OF AFLATOXIN B1.
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.
CC -1- TISSUE SPECIFICITY: LIVER.
CC -1- INDUCTION: BY 3-METHYLCOLANTHRENE (3MC).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M63788; AAA37068.1; -
DR EMBL; M27906; AAA37084.1; -
DR PIR; A33293; AAA37084.1; -
DR PIR; S13884; S13884.
DR InterPro; IPR001128; -
DR InterPro; IPR002401; -
DR Pfam; PF00067; P450.1.
DR PRINTS; PR00385; P450.
DR PRINTS; PR00463; EP450I.
DR PROSITE; PS00086; CYTOCHROME_P450.1.
DR Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW Microsome; Endoplasmic reticulum.
FT BINDING 439 439 HEME (BY SIMILARITY).
FT CONFLICT 417 417 S->L (IN REF. 2).
SO SEQUENCE 494 AA; 57387 MW; 16816672AAAB112B CRC64;

Query Match 11.3%; Score 302; DB 1; Length 494;
Best Local Similarity 24.2%; Pred. No. 2,4e-12;
Matches 128; Conservative 85; Mismatches 228; Indels 88; Gaps 20;

QY 12 LLLPSSLILFLILKRRNRKTRFNLPRKSGWPLIGETIGYIKPYTATTLGDFMOOH 71
DB 6 MLVAVVLTCLSVMTIMSVQRRLRKMPGPPTLPFIQNL----ELDTKRYDCLSKM 61
QY 72 VSKYGIYRSLNLFGEPTIV--SADAGLNRTILIONGRLEFECSSYPSISGILGKMSMLVLY 129
DB 62 RERYGPVTTILGPRPAVLMGYDAVKAALIDQAE-ELSDRGEQAFDFWFEKGIQVSS 120
QY 130 GDMHRDMRSISLNFSLH-----ARLRTI-----LLKDVRRHTL--FVIDSMOON 171
DB 121 GERAKQLRRFSIATLRDGFGRGIEERTIETISFLIALDNTGATIDPTFYMSRTVSN 180
QY 172 SIFSA-----QDEAKKTFNLMAKHMSPDGEETEQLKKEYTFMKGYVASAPLNLPG 225
DB 181 VISSIVFGNREYDKEF-LSLIGMIRSFOFMSTSTQLEFEMFYSVK-----HLPG 232
QY 226 TAYKALQSRATILKFERKME--RKLDIKEEDQ--EEEEKTFDEAEMSKSD-HVRKQ 280
DB 233 -CQHDAKEMOGLEDFIARKKEENORTLDPSNPRDFIDSFILRKQEEKKNPTQOFHMR-- 289
QY 281 RTDDDLGWLKHSNLSLEQILDLILSLFAGHETSVAILAIFLQACPKAVAEELREE 340
DB 290 -----NLMTLTNLNFFAGTETVSTTTTRGFLLLMKYPHIAKMHHE 330

```

```

QY 341 HLEIARAKKEIGES-ELAWDDYKKKDFTCYINETLRIGNVYRF-LHRKALKDVRKRGYD 398
DB 331 IDQV-----IGRNQPKYEDHLKMPYTEAVIETIQRFVDVVPVLGLPSTTKDKIFRDL 384
QY 399 IPGKRVLPVISAVALDNSRYDQPNLFNPMRMQOONNCASSSGSFSFTGNNVMPFGCG 458
DB 385 IPKGTDFPVLLSSVLDKDPKFSNPNDFNPQH-----SDDKQGFKK-SNAFMPFSVG 435
QY 459 PRLGSELAKLEMAVFTIHLVLKFNWELAE----DDQPAFPVDFP 502
DB 436 KRYCGESLAKMELFIFFTITMGNCFKSPQAPQDIDVTPOYFSFAALP 484

```

Search completed: October 17, 2001, 22:14:09  
Job time: 363 sec





```

Db 121 GKWSMLVGVGMHMRDMSISLNFLSHARLRTLLKDYERHTLFVLDMSQNSIFSADDEA 180
QY 181 KKFENMLAKHIMSMDPGEETEOLKKEYVTFMKGVSAPLNPGTAVHKAQSRATILK 240
Db 181 KKFENMLAKHIMSMDPGEETEOLKKEYVTFMKGVSAPLNPGTAVHKAQSRATILK 240
QY 241 FIERKMEERKLDIKEEDOEVEEVKTEDEAEKSDHVKRQKOTDDLLGWLKHSNLSREQ 300
Db 241 FIERKMEERKLDIKEEDOEVEEVKTEDEAEKSDHVKRQKOTDDLLGWLKHSNLSREQ 300
QY 301 ILDLILSLFAGHETSSVAIALAIFLQACPKAVEELREHELEIARAKKEGSELMND 360
Db 301 ILDLILSLFAGHETSSVAIALAIFLQACPKAVEELREHELEIARAKKEGSELMND 360
QY 361 YKKMDFQCVINETRLGNNVRFLLHRAKAKDVRKGYDIPSGMKVLPVISAHLNDSRYD 420
Db 361 YKKMDFQCVINETRLGNNVRFLLHRAKAKDVRKGYDIPSGMKVLPVISAHLNDSRYD 420
QY 421 QPNLFNPMRMOQONNGASSSGSGSFSTWGNMYPFGGPRLCAGSELAKLEMAVFIHHLV 480
Db 421 QPNLFNPMRMOQONNGASSSGSGSFSTWGNMYPFGGPRLCAGSELAKLEMAVFIHHLV 480
QY 481 LKFNWELAEDDQPPAFPEVDFPNGLPIRVSRIL 513
Db 481 LKFNWELAEDDQPPAFPEVDFPNGLPIRVSRIL 513

```

```

RESULT 2
Q9SCQ9 PRELIMINARY: PRT: 513 AA.
AC Q9SCQ9
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE STEROD 22-ALPHA-HYDROXYLASE (DWF4).
GN T345.40.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicaceae; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bioecker H., Mewes H.W., Lemcke K., Mayer K.F.X., Quetier F.,
RA Salanoubat M.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC EMBL: AL132979; CAB62435.1; -.
DR InterPro: IPR001128; -.
DR Pfam: PF00067; P450.1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
KM Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 513 AA; 58867 MW; B1639BD9A5D7C93 CRC64;

```

```

Query Match 99.9%; Score 2677; DB 10; Length 513;
Best Local Similarity 99.8%; Pred. No. 1.2e-182;
Matches 512; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MFEETHTLPLLLPSLSLLFLILKRRNKRTRNLPFGKSGWPLFGTGYLKPRT 60
Db 1 MFEETHTLPLLLPSLSLLFLILKRRNKRTRNLPFGKSGWPLFGTGYLKPRT 60
QY 61 ATTGADPMQOVSKYKRYRNLGEPRTIVSADAGLNRTIQLNGRGLFECYPRISIGIL 120
Db 61 ATTGADPMQOVSKYKRYRNLGEPRTIVSADAGLNRTIQLNGRGLFECYPRISIGIL 120

```

```

QY 121 GKWSMLVGVGMHMRDMSISLNFLSHARLRTLLKDYERHTLFVLDMSQNSIFSADDEA 180
Db 121 GKWSMLVGVGMHMRDMSISLNFLSHARLRTLLKDYERHTLFVLDMSQNSIFSADDEA 180
QY 181 KKFENMLAKHIMSMDPGEETEOLKKEYVTFMKGVSAPLNPGTAVHKAQSRATILK 240
Db 181 KKFENMLAKHIMSMDPGEETEOLKKEYVTFMKGVSAPLNPGTAVHKAQSRATILK 240
QY 241 FIERKMEERKLDIKEEDOEVEEVKTEDEAEKSDHVKRQKOTDDLLGWLKHSNLSREQ 300
Db 241 FIERKMEERKLDIKEEDOEVEEVKTEDEAEKSDHVKRQKOTDDLLGWLKHSNLSREQ 300
QY 301 ILDLILSLFAGHETSSVAIALAIFLQACPKAVEELREHELEIARAKKEGSELMND 360
Db 301 ILDLILSLFAGHETSSVAIALAIFLQACPKAVEELREHELEIARAKKEGSELMND 360
QY 361 YKKMDFQCVINETRLGNNVRFLLHRAKAKDVRKGYDIPSGMKVLPVISAHLNDSRYD 420
Db 361 YKKMDFQCVINETRLGNNVRFLLHRAKAKDVRKGYDIPSGMKVLPVISAHLNDSRYD 420
QY 421 QPNLFNPMRMOQONNGASSSGSGSFSTWGNMYPFGGPRLCAGSELAKLEMAVFIHHLV 480
Db 421 QPNLFNPMRMOQONNGASSSGSGSFSTWGNMYPFGGPRLCAGSELAKLEMAVFIHHLV 480
QY 481 LKFNWELAEDDQPPAFPEVDFPNGLPIRVSRIL 513
Db 481 LKFNWELAEDDQPPAFPEVDFPNGLPIRVSRIL 513

```

```

RESULT 3
Q9LKH7 PRELIMINARY: PRT: 474 AA.
AC Q9LKH7
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE CYTOCHROME P450.
GN CYP7C.
OS Phaseolus aureus (Mung bean) (Vigna radiata).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fabales; Fabaceae; Papilionoideae; Vigna.
OX NCBI_TaxID=3916;
RN [1]
RP SEQUENCE FROM N.A.
RA Yang M.T., Chen Y.M.;
RA Yang M.T., Chen Y.M.;
RL "Cloning and sequencing of a Vigna radiata cDNA encoding cytochrome
RT P450."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL: AF279252; AF89209.1; -.
DR InterPro: IPR001128; -.
DR Pfam: PF00067; P450.2.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
KM Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome;
KM Monooxygenase; Oxidoreductase.
SQ SEQUENCE 474 AA; 54037 MW; A2542A809C5BACD CRC64;

```

```

Query Match 35.5%; Score 951; DB 10; Length 474;
Best Local Similarity 40.1%; Pred. No. 8.2e-60;
Matches 203; Conservative 97; Mismatches 162; Indels 44; Gaps 10;

```

```

QY 12 LLLPSLSLLF-----LILKRRNKRTRNLPFGKSGWPLFGTGYLKPRTATTLGD 66
Db 1 MVSILPTL-LILFAASAIAIFLHRAFSRRKRLPPGSGYLPIETIQLTISAVKSSNPEP 58

```



DR PROSITE: PS00037; MYB\_1; UNKNOWN\_1.  
 KW Electron transport; Endoplasmic reticulum; Heme; Hypothetical protein;  
 KM Membrane; Microsome; Monooxygenase; Oxidoreductase.  
 SQ SEQUENCE 524 AA; 59389 MW; 550578908BDDF272 CRC64;

Query Match 32.2%; Score 864; DB 10; Length 524;  
 Best Local Similarity 35.4%; Pred. No. 1,5e-53;  
 Matches 186; Conservative 96; Mismatches 165; Indels 78; Gaps 10;

QY 24 FLIL-----LKRNRKRTFN-----LPPGSGVFLGFTIGYL--- 56  
 DB 30 FLVLTAGILRPWLMFRLNRKSTKDGDEEDNEKKKGKMGSLGAVIGETLNFILGCG 89  
 QY 57 ---KPYATTLGDMQOVHSKYGKIRSNLFGCEPTVSADAGLNFILQNGRLPECYSP 113  
 DB 90 YSNRPV-----FMDKRSKLYGKVFKNITIGTPIITSTDAEVKVVLONGNFTVPAYP 143  
 QY 114 RSIGGILGKMSLVLDGMDRMSISLNLISHARLTILIKDVERHTLFLVDSMOQNSI 173  
 DB 144 KSTIELLENSILSINGPHQRRLHTLIGAFRLSPHLKDRITRDEASVLTLASWAQLPL 203  
 QY 174 FSADEAKKFTFNIMAKHNSMDGEEETDQKKKYTFMKGVVSAPLNLPGTAVAKALQ 233  
 DB 204 VHVDEIKKMTFELLYKVMSTSP-EDMNLKLEFEFIFGLICIPKPGTRLKYSLK 262  
 QY 224 SRATLKEIERKMEERKLDIKEEDQEEVEYKT-----EDEAENSKSDHVRKQRTDDL 287  
 DB 263 AKERLIKVKVVEERQVAMTTSPANDVVDLLRDGDSKQSPDSFVSGK----- 315  
 QY 288 GWLKHSMSTEQILDLISLFGHETSSVALAIFLQACPKAVELEBEHELEARA 347  
 DB 316 -----IVEMKIPGEETMPTAMTLAVKFLSDNPVALAKLEENEMKRR 358  
 QY 348 KKEGESELMNDYKKKMFQCVINETRLGNVYRFLHRKRLKDRYGYIPSGKVLPL 407  
 DB 359 KLEIGE-EYKWTDTWMSLFTQNVINETLRMANITLNGWRKALKDELIGYILPGWCVLA 417  
 QY 408 VISAVHLNDSRYDQNLNFPWRMOQNGASSSGSFSTGNNMPPRGCGPRLCAGSEL 467  
 DB 418 SFISVHMEDELYDNPQYQDPWRMDRINGSANSIC-----FTPPGQGRILCPGSEL 468  
 QY 468 AKLEMAVFHHLVLFKFMELAEDDQPAFPVDPNPGLPIVRSRI 512  
 DB 469 SKLEISIFLHLVTRYSW-TAEDEIVSFPTVKMKRRLPIVATV 512

RESULT 6  
 ID Q9M066 PRELIMINARY; PRT; 457 AA.  
 AC Q9M066;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)  
 DE CYTOCHROME P450 LIKE PROTEIN.  
 GN A14G36380.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP PARTIAL SEQUENCE FROM N.A.  
 RA Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,  
 RA Meves H.W., Lemcke K., Meyer K.F.X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +  
 CC OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY

CC SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 DR EMBL: AL161589; CAB80304.1; -  
 DR InterPro: IPR001005; -  
 DR InterPro: IPR001128; -  
 DR Pfam: PF00067; P450; 1.  
 DR PRINTS: PR00385; P450.  
 DR PROSITE: PS00086; CYTOCHROME\_P450;  
 DR PROSITE: PS00037; MYB\_1; UNKNOWN\_1.  
 KW Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome;  
 KM Monooxygenase; Oxidoreductase.  
 SQ SEQUENCE 457 AA; 51607 MW; BD6A148E4EACD2 CRC64;

Query Match 32.1%; Score 860; DB 10; Length 457;  
 Best Local Similarity 36.8%; Pred. No. 2,4e-53;  
 Matches 179; Conservative 93; Mismatches 160; Indels 54; Gaps 8;

QY 39 LPPEKSGVFLGFTIGYL-----KPYATTLGDMQOVHSKYGKIRSNLFGCEPTVSA 92  
 DB 2 IPNSGLGWPVIGETLNFILFACGYSSRPV-----FMDKRSKLYGKVFKNITIGTPIIIST 55  
 QY 93 DAGNRLFTLQMEGRIFECSPRSITGILGKMSLVLDGMDRMSISLNLISHARLTIL 152  
 DB 56 DAEVKKVVLQNHGNTFVPAYPKSTIELLENSILSINGPHQRRLHTLIGAFRLSPHLKDR 115  
 QY 153 LKDVERTLFLVDSMOQNSIFSADEAKKFTFNIMAKHNSMDGEEETDQKKKYTF 212  
 DB 116 ITROLEASVLTLASWAQLPLVHVODEIKKMTFELLYKVMSTSG-EDMNLKLEFEF 174  
 QY 213 MKGVVSAPLNLPGTAVAKALOSRATILKIERKMEERKLDIKEEDQEEVEYKT-----E 266  
 DB 175 IKGLICIPKPGTRLKYSKAKERLIKMKVVEERQVAMTTSPANDVVDVLLRDGCD 234  
 QY 267 DEAEKSKSDHVRKQRTDDDLGWLKHSNLTSEQLDLISLFGHETSSVALAIAFF 326  
 DB 235 SEKQSPDEVSFGK-----IVEMKIPGEETMPTAMTLAVKFLSDNPVALAKLEENEMKRR 270  
 QY 327 LQACPKAVELEBEHELEARAKKESELMNDYKKMFQCVINETRLGNVYRFLHR 386  
 DB 271 LSDNPVALAKLEENEMKRRKLEIGE-EYKWTDTWMSLFTQNVINETLRMANITLNGWR 329  
 QY 387 KALDVRKGYDIPSGKVLVVISAVHLNDSRYDQNLNFPWRMOQNGASSSGSFS 446  
 DB 330 KALKDVEIKGYLIPGWCVLASFISVHMEDELYDNPQYQDPWRMDRINGSANSIC----- 385  
 QY 447 TWGNMYPFGGGRPLCAGSELAKLEMAVFHHLVLFKFMELAEDDQPAFPVDPNPGLP 506  
 DB 386 -----FTPPGQGRILCPGLSELKLEISIFLHLVTRYSW-TAEDEIVSFPTVKMKRRLP 439  
 QY 507 IRVRSRI 512  
 DB 440 IRVATV 445

RESULT 7  
 ID Q9LN73 PRELIMINARY; PRT; 478 AA.  
 AC Q9LN73;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)  
 DE T12C24.27.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Chao Q., Brooks S., Buehler E., Johnson-Hopson C., Khan S., Kim C.,  
 RA Shinn P., Altafi H., Bai Q., Chin C., Chiu J., Choi E., Conn L.,  
 RA Conway A., Gonzales A., Hansen N., Howng B., Koo T., Lam B., Lee J.,

RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,  
 RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thayerl A.,  
 RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,  
 RA Ecker J.R.,  
 RT "Genomic sequence for Arabidopsis thaliana BAC T12C24 from chromosome  
 RT 1."  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [12]  
 RP SEQUENCE FROM N.A.  
 RA Ecker J.R.,  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [13]  
 RP SEQUENCE FROM N.A.  
 RA Ecker J.R.,  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 RN [14]  
 RP SEQUENCE FROM N.A.  
 RA Cheuk R., Shin P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,  
 RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chlou J., Choi E.,  
 RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,  
 RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,  
 RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,  
 RA Theologis A., Ecker J.,  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 DR EMBL: AC025417; AAF88087.1; -.  
 DR InterPro: IPR001128; -.  
 DR Pfam: PF00067; P450.1.  
 DR PRINTS: PR00385; P450.  
 DR PROSITE: PS00086; CYTOCHROME\_P450; UNKNOWN\_1.  
 KW Heme: Monooxygenase; Oxidoreductase.  
 SQ SEQUENCE 478 AA; 55057 MW; F67A9B9F2FD28558 CRC64;

Query Match 30.7%; Score 822; DB 10; Length 478;  
 Best Local Similarity 35.6%; Pred. No. 1.3e-50;  
 Matches 181; Conservative 95; Mismatches 187; Indels 46; Gaps 8;

QY 12 LLLPSSLILFLILKRRNRKTRFNLPCKSGMPFLGFTIGYAKPYATTATLGDPMOQH 71  
 DB 4 LLIWVSLILSITHVWVSRWPKCKGKLPKSGMPLGSGIOFPKPKTSIDIPPIKER 63  
 QY 72 VSKYGIKRSNLFGEPTVSADAGLNRFILNCEGRLEFCSYPRISGILGKMSMLVLGD 131  
 DB 64 VKKGPRIKTLNKGPRVIVSTDADLSYFVNOEGRCFQSWTPDTTHLFGKKNVSLHGF 123  
 QY 132 MHRDRSISLNFSLHARLRTLLKDVREHTLFVLDSDMOONSIFSAODEAKKFTFMMAKH 191  
 DB 124 MYKYLKNNVLTFLGHDLK -MLPQVEMTANKRLELWNSODVELKDAVSIIDPLTKK 182  
 QY 192 IMSDGEELFQKKEVYFMKGVSAPLNPGTAHYKALOSRATILKFERKKEERKL 251  
 DB 183 LISHP -KSSSENLANVAFIOLGISFPFPIPGTAHYKCLQGRAKAKMLNMLOERR- 240  
 QY 252 DIKEEDQEEBEVKTDEAKMSKSDHYRKQRD -DDLLGWVLKHSNLSLEQI-LDLIISL 308  
 DB 241 -----EMPRKNPSDFEYVIEEIOKEGTITLLEALDLMLVL 277  
 QY 309 LFAGHETSSVAIALAIFLQACPKAVEELREHELIARAKLGESELNWDYKMDFTQ 368  
 DB 278 LFASSETTSLATLAKFLSDOPEVLKRLTEHEHTLR -NRDADASGLTWETKMTYTF 336  
 QY 369 CVINETLGLGNVRFLLHKKALKDVRYK-----GYDIPSGKVLFLVISAVHLNDSRYDP 422  
 DB 337 QFINETLANIAPAIFKALRDIKFEVDNTDTITPAGMAVWVCPAVHVNIPKMYDP 396  
 QY 423 NLFNMRQOONNGASSSGSFTWGNVMPFGGPRCLGASLAKLEMAVFIHHLVLK 482  
 DB 397 LVFNPSRME-----GSKVTNASHKHFNAFGGMRFCVGTDTTKLOMAFLHSLVTK 446  
 QY 483 FNMELAEDDQFPAPFVDFPNGLPIRVSR 511

DB 447 YRMEIRKGNITRTPGLOFPNGYHVKLHK 475  
 RESULT 8  
 ID Q9LY89 PRELIMINARY; PRT: 382 AA.  
 AC Q9LY89;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE HYPOTHETICAL 43.9 KDA PROTEIN.  
 GN F18022.190.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Beyer M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,  
 RA Rudd S., Lemcke K., Mayer K.F.X.,  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [12]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +  
 CC OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY  
 CC SIMILARITY).  
 DR EMBL: AL163817; CAB87779.1; -.  
 DR InterPro: IPR001128; -.  
 DR Pfam: PF00067; P450.1.  
 DR PROSITE: PS00086; CYTOCHROME\_P450; UNKNOWN\_1.  
 KW Electron transport; Endoplasmic reticulum; Hypothetical protein;  
 KW Membrane; Microsome.  
 SQ SEQUENCE 382 AA; 43888 MW; 1BC5685AB73BAE30 CRC64;

Query Match 30.4%; Score 815; DB 10; Length 382;  
 Best Local Similarity 34.2%; Pred. No. 2.9e-50;  
 Matches 159; Conservative 87; Mismatches 127; Indels 92; Gaps 4;

QY 45 GMPFLGFTIGYAKPYATTATLGDPMOQHSYKGIKRSNLFGEPTVSADAGLNRFILONE 104  
 DB 2 GMPFLGFTISFPRRSISIGFTLOORSRYGKVKRSNCGGKAVVSCDQETNMFILONE 61  
 QY 105 GLPFCSPRSISGILGKMSMLVLGDMHRDRSISLNFSLHARLRTLLKDVREHTLFV 164  
 DB 62 GKLFISDPKAMHDLTGKYSLLATGELIHRKLNVIISFINITKSKPDLHGAENLSISI 121  
 QY 165 LDSMOONSIFSAODEAKKFTFMMAKHIMSMDGEEETQKKEVYFMKGVSAPLN 224  
 DB 122 LKSWNCREVEFRHKVKTFTLSVNVQLLSIKPEDPARLYVLDDLSYWKGISIDPLP 181  
 QY 225 GTAHYKALOSRATILKFERKKEERKLDIKEEDQEEBEVKTDEAKMSKSDHYRKQRD 284  
 DB 182 GTGYNAIKVRSNRNIHONALIEDNMNNAIREDPFLDSIISND----- 224  
 QY 285 DLGGLVNLKHSNLSLEQIIDLILSLFAGHETSSVAIALAIFLQACPKAVEELREHELEI 344  
 DB 225 -----BEHMAI 230  
 QY 345 ARAKLEGESELNWDYKMDFTQCVINETLGLGNVRFLLHKKALKDVRYKYDIPSGWK 404  
 DB 231 -RAKKGDEL -LNMEDYKMEFTQCVISALRCGNIVTVIHKRKAHDIKFNFYVIPKWK 288  
 QY 405 VLPVISAHLNDSRYDPPLFNPWRMOOONNGASSSGSFTWGNVMPFGGPRCLG 464  
 DB 289 VFIFTAAHLDSLHNDPEFNPMKMTKT-----AFQGVAVCPG 329  
 QY 465 SELAKLEMAVFIHHLVLKFNWELAEDDQFPAPFVDFPNGLPIRV 509

Db 330 GELCKLQIAFFLHLVLSTYRKIKSDENPIAHVPEFRGMLLET 374

RESULT 9

09LIC5 PRELIMINARY; PRT: 464 AA.

AC 09LIC5; (Tremblrel. 15, Created)

DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)

DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)

DE CYTOCHROME P450-LIKE PROTEIN.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;

OC Brassicales; Brassicaceae; Arabidopsis.

NCBI\_TaxID=3702;

RP SEQUENCE FROM N.A.

RC STRAIN=COLUMBIA;

RA Kaneo T., Kato T., Nakamura Y., Asamizu E., Tabata S.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

PP SEQUENCE FROM N.A.

RC STRAIN=COLUMBIA;

RA Kaneo T., Kato T., Nakamura Y., Asamizu E., Tabata S.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

PP SEQUENCE FROM N.A.

RC STRAIN=COLUMBIA;

RA Kaneo T., Kato T., Nakamura Y., Asamizu E., Tabata S.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

PP SEQUENCE FROM N.A.

RC STRAIN=COLUMBIA;

RA Kaneo T., Kato T., Nakamura Y., Asamizu E., Tabata S.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

PP SEQUENCE FROM N.A.

RC STRAIN=COLUMBIA;

RA Kaneo T., Kato T., Nakamura Y., Asamizu E., Tabata S.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

PP SEQUENCE FROM N.A.

RC STRAIN=COLUMBIA;

RA Kaneo T., Kato T., Nakamura Y., Asamizu E., Tabata S.;

Qy 457 GGPRLCAGSELAKEMAVFIHILVLE 483

Db 437 GGRLCPGLDLARLETSYFLHILVTRF 463

RESULT 10

09FMA5 PRELIMINARY; PRT: 465 AA.

AC 09FMA5; (Tremblrel. 16, Created)

DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)

DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)

DE CYTOCHROME P450.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;

OC Brassicales; Brassicaceae; Arabidopsis.

NCBI\_TaxID=3702;

RP SEQUENCE FROM N.A.

RC STRAIN=COLUMBIA;

RA Kaneo T., Kaneo T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [1]

PP SEQUENCE FROM N.A.

RC STRAIN=COLUMBIA;

RA Kaneo T., Kaneo T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [1]

PP SEQUENCE FROM N.A.

RC STRAIN=COLUMBIA;

RA Kaneo T., Kaneo T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [1]

PP SEQUENCE FROM N.A.

RC STRAIN=COLUMBIA;

RA Kaneo T., Kaneo T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [1]

PP SEQUENCE FROM N.A.

RC STRAIN=COLUMBIA;

RA Kaneo T., Kaneo T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [1]

PP SEQUENCE FROM N.A.

Query Match 27.1%; Score 726; DB 10; Length 465;

Best Local Similarity 32.3%; Pred. No. 8.4e-44;

Matches 166; Conservative 91; Mismatches 181; Indels 76; Gaps 10;

Db 11 LLIIVSLCSALL-----RNQMKYTKNGLPFGTGMPIFETETFLKQGP-----NMR 59

Qy 70 QHVSRYGKRYNSLGEPTIVSADAGLNRFLQNEGRLFECSYPRISIGILGKMSMLV 129

Db 60 NQRLRYGKRYNSLGEPTIVSADAGLNRFLQNEGRLFECSYPRISIGILGKMSMLV 119

Qy 130 GDMRDRKRSILNLSHARLRTILKDYERHTLVLDSDMOONSIFSADDEKAKTFP---- 185

Db 120 GSSHRKMGSLSLISSTMDMDHILPKVDHFRSRYLDQWNELEVIDIQDKYHMAFLSSL 179

Qy 186 -----NLMAKHIMGMDGEEETEQLEKKEYVTFMKGVVAPLNLGTAHAKLQSRATILK 240

Db 180 TQINGNLKRPV-----EEFKTAFKLVYGLTSLVPLDPLGTNRRCIOQRNNIDR 229

Qy 241 FIERKMEERKLDKEEDQEEVEKTEDEAKMSKSDHVRKQRTDDLLGLWGLKHSN---L 296

Db 230 LLRELQGERR-----DSGE-----TFIDMLGYLKKKGNRYPL 262

Qy 297 STEQIDLLILSLFAGHETSSVATAIAIFLQACPKAVEELREHLELARKKKEGSEL 356

Db 263 TDEIRQOVVITLISGYETVSTSMMLKYLHDPKALQELRAHLEFRERKRO--DEPL 320

Qy 357 NMDYKMDKTCQVINEFLRGVNVREFLHRKALDVRKRGKIDIPSGMKVPLVISAHLND 416

Db 321 GLEVKMKKFTRAVIYETSLRATIVNGVLKRTTDLDELNGILKGRIRIYYTTEINYDA 380

Qy 417 SRYDQPLNFPMWQOONNGASSSGSFSFTWGNVYMPFGGPRLCAGSELAKEMAVFI 476

Db 381 NLVEDPLIFNPMWQOONNGASSSGSFSFTWGNVYMPFGGPRLCAGSELAKEMAVFI 429

Qy 477 HHVLKFNWELAEEDOPFAPFVDFPGLPIRVS 510

Db 430 HYFVTRWELIGDELAVFPRVAPKGFHLIRIS 463





```

RESULT 13
065624 PRELIMINARY: PRT: 457 AA.
AC 065624:
DT 01-AUG-1998 (TREMblrel. 07, Created)
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE CYTOCHROME P450.
CN T18B16.200 OR AT4G19230.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OC NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Beyer M., Benes V., Rechmann S., Borkova D., Ansoerge W., Bancroft I.,
RA Mewes H.W., Mayer K., Schueller C.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Benes V., Rechmann S., Borkova D., Ansoerge W., Mewes H.W., Lemcke K.,
RA Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 131-457 FROM N.A.
RA Van Der Schueren J., Chuang Y.J., Aert R., Defoor E., Robben J.,
RA Volckert G., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AL021687; CAAT6713.1; -
DR EMBL; AL161550; CAB78925.1; -
DR HSSP; P33006; 1CPT.
DR InterPro; IPR001128; -
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ
Sequence 457 AA; 52436 MW; 6105PB7C1B1E4F07 CRC64;

Query Match 23.6% Score 632; DB 10; Length 457;
Best Local Similarity 31.88; Pred. No. 4e-37;
Matches 151; Conservative 92; Mismatches 176; Indels 56; Gaps 12;

QY 13 LLLPSSLFLFLILKRRNRKTRFNLPPGKSGWFLGTTGYLKPYATTTLGDFMOQH 72
DB 10 LFAGSLFLYFLRCLISORFGSSKLPPLPGTMGWYGET---FQLYSQDP-NVFPQSKQ 65

QY 73 SKYGIYNSNLFGEPTIYSAAGLNRFILONEGRLEFESYSTRSIGILGKSMVLVGD 132
DB 66 KRYGSVFTHVLGDCVWISSPEAKFVLVTKSHLFKTFPASKERMGKQAIFEHGDY 125

QY 133 HRDRSISLNFSLHARLITLLKDVERTLFLVLDWMQNSIFSAODEAKKFTFNLMAHI 192
DB 126 HAKLRKLVLRPFMEPSTIN-WPDIESTADQSLNSW-EGTIMINTYOEKKTTFENVALI 183

QY 193 MSMDPGEET---DOLKEVYTFMGKGVSAFLNPGTAVHAKLOSRAITLFIERKMEER 249
DB 184 F---GKREVLRYDLKRCYILIEKGYNSMVPNLPGTLFHKSMARKELSDILARKILSR 239

QY 250 KLDIKEEDQEEVEKTEDEAEMSKSDHVRKQRTDDLLG-VLXHSNISTEQILDLISL 308
DB 240 R-----QNGSSH-----NDLGSFGKDEELTDQIADNITIGV 272

QY 309 LFAHETSSVALAIFLOACPKAVELREHLEIARAKKELGESLNMWDYKMDPTQ 368
DB 273 IFAARDTTASVSMILKYLAEINPVLEAVTEEQMAI-RKREGEBS-LTWGDTKKMPLTS 330

```

```

QY 369 CVINETLRGNVRELRHKKALKDVRKCYDIPSGWKVLPVISAVHLNDSRDOPULFPMW 428
DB 331 RVIGETLRVASTLSFTFEAVEDVEYEGTLLPKGKVLPRLNRTHSHADITSNCKEFPDS 390

QY 429 RMQOONNCGSSSGSFSTGNNVMPFGGPRLCAGSELAKLENAVFLTHLVLFK 483
DB 391 RFE-----VAPKRNTEFMFPCNGTHSCGNELAKLEMSIMHLLTTKV 432

RESULT 14
09FI38 PRELIMINARY: PRT: 518 AA.
AC 09FI38:
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE CYTOCHROME P450-LIKE PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OC NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-COLUMBIA;
RX MEDLINE=99397451; PubMed=10470850;
RA Kaneo T., Katoh T., Sato S., Nakamura Y., Asamizu E., Kotani H.,
RA Miyajima N., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. IX.
RT Sequence features of the regions of 1,011,550 bp covered by seventeen
RT P1 and TAC clones."
RL DNA Res. 6:183-195(1999).
DR EMBL; AB017064; BAB11064.1; -
SQ
Sequence 518 AA; 58877 MW; CCA3F74F4301563E CRC64;

Query Match 23.5% Score 631; DB 10; Length 518;
Best Local Similarity 30.5% Pred. No. 5.7e-37;
Matches 152; Conservative 101; Mismatches 190; Indels 56; Gaps 9;

QY 27 LKRRNRKTRFNLPPGKSGWFLGTTGYLKPYATTTLGDFMOQHSYKGIYNSNLFGE 86
DB 64 LYRWSNPKRCNGKLPFGSGMLPIIGTCDFEPHGLYEISPPVKRMKLKYLGLFRNIGS 123

QY 87 PTIYSAOAGLNRFILONEGRLEFESYSTRSIGILGKSMVLVGDMDHMSISLNFSLH 146
DB 124 NTVVLEPDIITFEVFRQDNKSFVSYPAYVKPRGEKENVLKHGTHVHVKQISLOHLGS 183

QY 147 ARLRTILKDVERTLFLVLDWMQNSIFSAODEAKKFTF-NLMAKHMSMDPGEETPOL 205
DB 184 EALKKMGIDRYVLEHLSKANGSFDKAVEESVIMAHLPKTIISNLKPEQAT--L 241

QY 206 KKEVYTFMGKGVSAFLNPG-TAVHAKLOSRAITLFIERKMEERK-----LDIKE 255
DB 242 VDNIMALGSEWFGSPRLKLTLLISITKYFIARVALQYIKVDVTRRKASREMGDFLDMV 301

QY 256 EDOEEVEKTEDEAEMSKSDHVRKQRTDDLLGWVLKHSNISTEQILDLISLFAHET 315
DB 302 EGEKEEDYFNES-----AINLITAILVAKES 330

QY 316 SVAVALAIFLOACPKAVELREHLEIARAKKELGESLNMWDYK-KMDPTQCVINET 374
DB 331 TTSVTSIAIKFLAEHKKALAELEKREHALLDNRRNGK-AQSVWEYRQMTFTMNVINET 389

QY 375 LRLGNVRELRHKKALKDVRKCYDIPSGWKVLPVISAVHLNDSKRDOPNLFPMRQOON 434
DB 390 LRMANMAPIMYRKAVNDVEIKGYTIPAGVIYAVIPPAVHFMDAITYENPLEFNPWRBEKE 449

QY 435 NGASSGSGSFSTGNNVMPFGGPRLCAGSELAKLENAVFLTHLVLFKFNELADDPF 494
DB 450 ---LRSGSKTF-----WFGGVGVQCVGAEFARLQISTIFHLLVTTTDSLQDSSEFT 499

```



